DOCUMENT RESUME

ED 165 989 SE 025 259

AUTHOR	Theobald, John Hamilton
TITLE	Attitudes and Achievement in Biology: An Investigation of the Effects of Student Attributes and Teaching Styles on Attitudes and Achievement in Secondary School Biology.
PUB DATE	77
NOTE	370p.; Ph.D. Dissertation, Monash University, Australia
EDRS PRICE	MF-$0.83 HC-$19.41 Plus Postage.
DESCRIPTORS	*Academic Achievement; Biology; Doctoral Theses; *Educational Research; *Instruction; Science Course Improvement Project; *Science Education; Secondary Education; Student Attitudes; *Student Characteristics; *Teaching Styles
IDENTIFIERS	Australia

ABSTRACT

Using a specially-constructed category system (CATS), biology lessons are analyzed for attribute-treatment interaction (ATI) and the data submitted to analysis. Relationships between fifteen individual predictors (general ability, four personality factors, two cognitive preferences, treatment, and seven first-order attribute-treatment interactions) and criterion variables (four attitude scales and two achievement tests) are studied using multiple regression analysis. The unique contribution of each term, controlling for all others, is tested. Among the results reported, "enjoyment in biology" is considered to be a function of a cautious, convergent outlook, and an intellectual disposition. General ability and an intellectual disposition lead to high scores on a memory task in biology, and on this criterion, class-centered teaching is considered better for students who are both open-minded and have a cognitive preference for rate learning, and individual-centered teaching is considered better for dogmatic students and those with a cognitive preference for principles. On an achievement test sampling higher cognitive processes, high general ability, open-mindedness, and a cautious personality are advantageous and individual-centered teaching is better than class-centered teaching for all but students with a highly intellectual disposition. (Author/MA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED165989

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ATTITUDES AND ACHIEVEMENT IN BIOLOGY

An investigation of the effects of student attributes and teaching styles on attitudes and achievement in secondary school biology.

John Hamilton Theobald

B.Sc., B.Ed. (Melb.)

Submitted in fulfilment of the requirements
for the Degree of Doctor of Philosophy in the
Faculty of Education at Monash University.

1977

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

John H. Theobald

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

SE 025 259

To

Alfred Austin Theobald
1892 - 1971

TABLE OF CONTENTS

Abstract vii

Declaration ix

Acknowledgements x

Tables xii

Figures xv

CHAPTER ONE : ORIGIN OF THE PROBLEM 1

CHAPTER TWO : TEACHING THE PROCESS OF INQUIRY 12

2.1 Biology as a process of inquiry into the living world 13

2.2 'Discovery' and 'inquiry' 14

2.3 Teaching strategies using the laboratory 15

2.4 Teaching strategies described as inductive or deductive 18

2.5 Teaching strategies described as guided discovery 21

2.6 Concluding observations 23

CHAPTER THREE: ATTRIBUTE-TREATMENT INTERACTION 26

3.1 Introduction 27

3.2 General ability and instructional treatment 34

3.3 Special abilities and instructional treatment 37

3.4 Personality attributes and instructional treatment 37

3.5 Cognitive style and instructional treatment 46

3.6 Prior achievement and instructional treatment 49

3.7 Conclusion 51

CHAPTER FOUR : THE DESIGN OF THE INVESTIGATION 53

4.1 Contrasting instructional treatments 54

4.2 Attribute-treatment interaction 55

4.3 Research in the classroom 56

4.4 Outline of the design 67

Setting 67

Treatment 69

Experimental population 70

Criterion tests 75

The analysis 76

CHAPTER FIVE : STUDENT CHARACTERISTICS: ATTRIBUTES 78

5.1 Intelligence and aptitudes 79

5.2 Personality characteristics 82

5.3 Measures of personality 87

5.4 The Omnibus Personality Inventory 90

5.5 Cognitive style: cognitive preferences 98

CHAPTER SIX : INSTRUCTIONAL TREATMENTS 109

6.1 Treatment variables: classrooms, courses, and teachers 110

6.2 Teacher variables: strategy and style 111

6.3 Contrasting styles 115

6.4 Observing classroom behaviour 119

6.5 Some category systems 120

6.6 The development of a category system 133

6.7 Recording and coding procedures 148

6.8 Validity and reliability 163

6.9 Validation and description of treatments 176

CHAPTER SEVEN: THE CRITERION TESTS 186

7.1 Introduction 187

7.2 Measures of cognitive outcomes 188

Construction of achievement tests 188

Sampling of process objectives 190

Estimates of reliability 191

7.3 Measures of non-cognitive outcomes 194

The role of the laboratory 194

The measurement of attitudes 196

7.4 The Biology Attitude Inventory scales 200

Description 200

Estimates of reliability 203

CHAPTER EIGHT: A MULTIPLE REGRESSION ANALYSIS 205

8.1 The intent of the analysis 206

8.2 Non-orthogonality of independent variables 210

8.3 Non-orthogonality between interactions and main effects 219

8.4 The appropriate error term 220

8.5 The attribute-treatment interactions 222

8.6 The main effects 225

CHAPTER NINE: INTERPRETATION OF FINDINGS 233

9.1 The final regression models 234

9.2 Statistical inference 239

9.3 The final trend equations 241

9.4 Views about biology learning (VABL) 243

9.5 Views about biology as a process (VABP) 245

9.6 Views about scientists (VAS) 251

9.7 Enjoyment of biology (EOB) 252
9.8 Achievement test AD1 (Diversity of living organisms) 253
9.9 Achievement test AI1 (Inter-relationships) 260

CHAPTER TEN: IMPLICATIONS OF THE RESEARCH 264
10.1 Summary of results 265
10.2 Implications 267
    Some disclaimers 267
    Implications 269
10.3 Suggestions for further research 273
10.4 Concluding observations 275

APPENDIX I: Unpublished instruments used in the present research 277
APPENDIX II: A category system for the analysis of teaching styles (CATS) 291
    Encoding sheet for CATS 299
APPENDIX III: Annotated transcript of sample lesson (Lesson 21) 300
    One-sixth sample coding of Lesson 21 327

Reference Notes 328
Reference List 329

ABSTRACT

The Web of Life biology course is an Australian adaptation of the Biological Sciences Curriculum Study materials, and is committed to the development of students' inquiry skills as well as knowledge of biology. Research on the advantages of teaching strategies utilizing the laboratory, inductive, and guided discovery methods is equivocal: this may be partly due to lack of validation of the instructional treatments and the use of inappropriate criterion tests, and partly because of attribute-treatment interaction. Although there have been few solidly established ATIs in educational research there is growing evidence that information-processing abilities may be more interactive than the traditionally researched abilities. Much ATI research has been criticized for its brevity and artificiality, and an argument is advanced for naturalistic experiments with minimal interference with school and class organization, course content, and instructional methods. Eight teachers were selected who were teaching the same course content towards the same agreed objectives by the same inquiry-centred teaching strategies, and who were known to be good teachers, but who were believed to possess contrasting natural teaching styles. Annotated transcripts of lessons were analyzed by a specially-constructed category system (CATS) and the data submitted to cluster analysis. Two contrasting groups were found, but contrary to expectations, with three in one group and five in the other. Multiple regression analysis enabled a parsimonious statement of relationships between fifteen individual predictors (general ability, four personality factors, two cognitive preferences, treatment, and seven first-order attribute-treatment interactions) and each criterion variable

(four attitude scales and two achievement tests). The unique contribution of each term, controlling for all others, was tested, first the interactions then the main effects, and those terms that did not significantly contribute to the explained variance were progressively eliminated. The total contribution of general ability was the last variable to be tested, acting as an independent predictor in its own right and as a covariate control for all other variables. General ability, an intellectual disposition, and open-mindedness contributed to favourable 'Views about biology learning'. General ability and an intellectual disposition also contributed to favourable 'Views about biology as a process', and on this criterion individual-centred teaching was better for dogmatic students with a cognitive preference for application, and class-centred teaching was better for open-minded students with a cognitive preference for theory. Very little of the total variance in scores on 'Views about scientists' was explained by the predictor variables. 'Enjoyment of biology' appears to be basically a function of a cautious, convergent outlook, and an intellectual disposition. General ability, and an intellectual disposition lead to high scores on a memory task in biology, and on this criterion class-centred teaching is better for students who are both open-minded and have a cognitive preference for rote learning, and individual-centred teaching is better for dogmatic students and those with a cognitive preference for principles. On an achievement test which samples higher cognitive processes, high general ability, open-mindedness, and a cautious personality are advantageous, and individual-centred teaching is better than class-centred teaching for all but students with a highly intellectual disposition.

DECLARATION

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma in any university and that, to the best of my knowledge and belief, the thesis contains no material previously published or written by another person, except when due reference is made in the text of the thesis.

John H. Theobald

John H. Theobald

ACKNOWLEDGEMENTS

My early experience in teaching convinced me that I was more successful with some students than with others on all kinds of criteria. Quite apart from their varied test results, some students appeared to respond more readily and some appeared to enjoy intellectual activity, but I did not seem to be able to influence others or make significant intellectual contact with them, no matter what I tried. All these students I thank for making me want to understand why.

I have always been encouraged by family, friends and colleagues. Many of them have helped in ways not now remembered, and others have been expected to make approving noises at the right time: their good sense and tact have been appreciated.

Barbara Fary and Elaine Atkinson assisted in the coding of lesson transcripts. The validity and reliability of the data so generated were central to the investigation, and both assistants will recognize the debt owed to their professionalism.

Leo West has been generous in the time he has given in association with the multiple regression analysis: the strategy of the analysis is his in conception. Lindsay Mackay acted as a temporary supervisor at the time of the processing of the various test results, and has maintained a consultative role in matters of research strategies in general and in the investigation of cognitive preferences in particular. I offer Dick White a sincere compliment in saying that he has been a demanding supervisor. He has set high standards of performance which have made the learning accompanying this investigation all the more satisfying.

The Faculty of Education provided support in purchasing published test materials, in the preparation and production of the unpublished tests, and in providing typing assistance for the first draft of the thesis. I thank those involved for the care they have taken: in particular Jacki Carton for typing most of the lesson transcripts, Reba Bloomfield for typing most of the draft, and Beverley Pocknee for typing Appendix III for inclusion in the thesis. Lynne O'Keefe typed the text of the thesis and her care and accuracy are warmly appreciated.

And finally, I pay tribute to the professionalism of all the teachers with whom I worked, and am only sorry that they must remain nameless. They readily agreed to participate in the investigation in the hope that our understanding of some aspects of teaching might be improved, and as evidence of good faith I offer them a copy of the summary of results and implications of the research.

J.H.T.

TABLES

4.4.1 t-TESTS FOR TREATMENT GROUPS ON THIRTEEN ATTRIBUTE VARIABLES 74

5.4.1 LOADINGS ON FACTORS EXTRACTED FROM TEN OPI SCALES 95

6.3.1 PROTOTYPES FOR SELECTION OF TEACHERS 116

6.6.1 DECISION LEVELS OF THE CATEGORY SYSTEM CATS 147

6.7.2 TOTAL CODED UNITS OF BEHAVIOUR FOR EACH TEACHER 154

6.7.3 SUMMARY OF CODED BEHAVIOUR: TEACHER 1 155

6.7.4 SUMMARY OF CODED BEHAVIOUR: TEACHER 2 156

6.7.5 SUMMARY OF CODED BEHAVIOUR: TEACHER 3 157

6.7.6 SUMMARY OF CODED BEHAVIOUR: TEACHER 4 158

6.7.7 SUMMARY OF CODED BEHAVIOUR: TEACHER 5 159

6.7.8 SUMMARY OF CODED BEHAVIOUR: TEACHER 6 160

6.7.9 SUMMARY OF CODED BEHAVIOUR: TEACHER 7 161

6.7.10 SUMMARY OF CODED BEHAVIOUR: TEACHER 8 162

6.8.1 ESTIMATES OF INTER-CODER AGREEMENT 170

6.8.2 ESTIMATES OF STABILITY OF TEACHER BEHAVIOUR 175

6.9.2 SIGNIFICANCE TESTS FOR DIFFERENCES BETWEEN INSTRUCTIONAL TREATMENTS BY GROSS CATEGORICAL DISTINCTIONS 182

6.9.3 SIGNIFICANCE TESTS FOR DIFFERENCES BETWEEN INSTRUCTIONAL TREATMENTS BY FINE CATEGORICAL DISTINCTIONS 183

7.2.1 CLASSIFICATION OF ACHIEVEMENT TEST ITEMS BY TAXONOMY CATEGORIES 191

7.2.2 ESTIMATES OF RELIABILITY FOR ACHIEVEMENT TESTS 192

7.2.3 QUARTILE ϕ AND FACILITY VALUES FOR ACHIEVEMENT TEST ITEMS 193

7.3.1 COMPARISONS OF SCALES OF PAI AND LUCAS' ADAPTATION 199

7.4.1 ESTIMATES OF RELIABILITY FOR ATTITUDE SCALES 203

8.5.1 VALUES OF R^2 AND F FOR (PERSONALITY FACTOR x TREATMENT) INTERACTIONS 223

8.5.2 VALUES OF R^2 AND F FOR (COGNITIVE PREFERENCE x TREATMENT) INTERACTIONS 224

8.5.3 VALUES OF R^2 AND F FOR (GENERAL ABILITY x TREATMENT) INTERACTIONS 224

8.6.1 VALUES OF R^2 AND F FOR MAIN EFFECTS: DEPENDENT VARIABLE 1 (VABL) 227

8.6.2 VALUES OF R^2 AND F FOR MAIN EFFECTS: DEPENDENT VARIABLE 2 (VABP) 228

8.6.3 VALUES OF R^2 AND F FOR MAIN EFFECTS: DEPENDENT VARIABLE 3 (VAS) 229

8.6.4 VALUES OF R^2 AND F FOR MAIN EFFECTS: DEPENDENT VARIABLE 4 (EOB) 230

8.6.5 VALUES OF R^2 AND F FOR MAIN EFFECTS: DEPENDENT VARIABLE 5 (AD1) 231

8.6.6 VALUES OF R^2 AND F FOR MAIN EFFECTS: DEPENDENT VARIABLE 6 (AI1) 232

9.1.1 F VALUE FOR EACH CRITERION VARIABLE, USING TREATMENT, SEVEN ATTRIBUTE VARIABLES, AND SEVEN ATTRIBUTE-TREATMENT INTERACTIONS AS INDIVIDUAL PREDICTORS, CONTROLLING FOR ALL OTHERS 235

9.1.2 PERCENTAGE OF VARIANCE IN EACH CRITERION VARIABLE EXPLAINED BY VARIOUS REGRESSION MODELS AND INDIVIDUAL PREDICTORS 236

9.1.3 RANKING OF CRITERION VARIABLES BY PARTICIPATING TEACHERS 237

9.2.1 MULTIPLE R, R^2, AND F VALUE FOR THE FINAL REGRESSION EQUATION FOR SIX CRITERION VARIABLES 240

9.3.1 INTERCORRELATIONS BETWEEN ATTRIBUTE VARIABLES 242

9.4.1 SELECTED STATISTICS FROM MULTIPLE REGRESSION FOR CRITERION VARIABLE VABL 243

9.5.1 SELECTED STATISTICS FROM MULTIPLE REGRESSION FOR CRITERION VARIABLE VABP 245

9.6.1 SELECTED STATISTICS FROM MULTIPLE REGRESSION FOR CRITERION VARIABLE VAS 251

9.7.1 SELECTED STATISTICS FROM MULTIPLE REGRESSION FOR CRITERION VARIABLE EOB 253

9.8.1 SELECTED STATISTICS FROM MULTIPLE REGRESSION FOR CRITERION VARIABLE AD1 254

9.9.1 SELECTED STATISTICS FROM MULTIPLE REGRESSION FOR CRITERION VARIABLE AI1 260

FIGURES

3.1.1 RELATIONSHIPS BETWEEN STUDENT ATTRIBUTES AND INSTRUCTIONAL TREATMENTS 29

6.7.1 PORTION OF ENCODED LESSON 152

6.9.1 CLUSTERING OF TEACHERS BY INVERTED FACTOR ANALYSIS 178

8.2.1 THE NON-ORTHOGONALITY OF GENERAL ABILITY, COGNITIVE PREFERENCES, AND PERSONALITY ATTRIBUTES 211

8.2.2 THE SEPARATION OF THE TOTAL CONTRIBUTION OF GENERAL ABILITY FROM OTHER ATTRIBUTE VARIABLES 212

8.2.3 REDEFINITION OF ORTHOGONAL ATTRIBUTE VARIABLES 213

8.2.4 THE ORTHOGONAL ATTRIBUTE-TREATMENT INTERACTION VARIABLES 216

9.3.2 INTERCORRELATIONS BETWEEN ATTRIBUTE VARIABLES 242

9.5.2 PLOT OF PERSONALITY FACTOR 4 (DOGMATISM) SCORES AGAINST VABP SCORES SHOWING THE PF4 x TREATMENT INTERACTION 248

9.5.3 PLOT OF COGNITIVE PREFERENCE SCALE Q - A SCORES AGAINST VABP SCORES SHOWING THE CP1 x TREATMENT INTERACTION 248

9.5.4 COMBINED PLOT OF PERSONALITY FACTOR 4 (DOGMATISM) SCORES AND COGNITIVE PREFERENCE 1 (SCALE Q - A) SCORES AGAINST VABP SCORES 249

9.8.2 PLOT OF PERSONALITY FACTOR 4 (DOGMATISM) SCORES AGAINST AD1 SCORES SHOWING THE PF4 x TREATMENT INTERACTION 257

9.8.3 PLOT OF COGNITIVE PREFERENCE SCALE P - M SCORES AGAINST AD1 SCORES SHOWING CP2 x TREATMENT INTERACTION 257

9.8.4 COMBINED PLOT OF PERSONALITY FACTOR 4 (DOGMATISM) SCORES AND COGNITIVE PREFERENCE 2 (SCALE P - M) SCORES AGAINST AD1 SCORES 259

9.9.2 PLOT OF PERSONALITY FACTOR 3 (NON-INTELLECTUALISM) SCORES AGAINST AI1 SCORES SHOWING THE PF3 x TREATMENT INTERACTION 262

CHAPTER 1

ORIGIN OF THE PROBLEM

Even at this short distance in time, the curriculum revision movement of the 1960's, particularly in the natural sciences, can be seen to have been a movement of the most profound educational significance. As with all great movements its origins are difficult to pinpoint with accuracy, although many of the contributing factors are clear enough.

In the decade following the Second World War the secondary school populations of the Western world grew enormously as a combined result of the baby-boom of the immediate post-war years and the tendency for children to stay on longer in secondary schools. The resources of school systems were strained to the limit and many critics felt that existing courses were not entirely appropriate for a less select secondary school population, which was now, in effect, all of the children of all of the people.

The orbiting of Sputnik vividly confirmed the industrial recovery of the USSR and its spectacular technological development since the Second World War. In the Cold War climate of national competition the spotlight was turned on the effectiveness of science education and, as Hurd (1961) points out, a number of critics felt that high schools in attempting to cater for all children had been 'watering down' their courses and that the "gifted and talented student was not being challenged to his capacity" (p.110). As Dershimer and Iannaccone (1973) have observed, the crystallisation of national concern which followed Sputnik produced massive federal support for curriculum innovation and the development of a wide range of curriculum materials. This move in America was followed by national curriculum projects in Britain, many funded by the Nuffield Foundation, and in many other countries.

The accelerated growth of science and technology has had other important consequences. The great breakthroughs in fields such as electron microscopy, nuclear physics, and space exploration have not only caused a vast accumulation of new scientific knowledge and have provided the technology and conveniences characteristic of industrialised society, but have added to the realisation of man's dependence on science and the social consequences of scientific discovery. Such considerations also added to the ground swell leading to science curriculum revision.

In this context there appeared an important book which both reflected the spirit of the times and gave a significant impetus and direction to the curriculum revisions. The Process of Education (Bruner, 1960) grew out of a national conference of American science educators at Woods Hole in 1959 and drew together a number of important themes which dominated the science curriculum revisions of the sixties. These themes were related to the structure of the subject matter of science, the nature of the learner, the nature of the teaching and learning process, and the technology of teaching.

The first of these themes is the notion of the structure of knowledge in science. As Gardner (1975c) says of the science curriculum developments of the sixties,

> it is possible to identify a common curricular ideology which has guided their construction. This ideology is related to a theory about the nature of academic disciplines, and is tied to the theory by a set of value judgments about the desirable relationships between these disciplines and their corresponding school subjects. (p.1)

Proponents of the 'structure-of-knowledge' theory such as Bruner (1960), Schwab (1964), and King and Brownell (1966) argue that school subjects should reflect faithfully the sciences from which they are drawn, not only in terms of the accuracy of the basic concepts but also in terms of the processes of inquiry that are used in these sciences.

The second theme, the nature of the learner, is derived largely from the work of Piaget. Piaget argues that intellectual development proceeds by physical maturation of the nervous system and interaction of the individual with his environment during which he is motivated by the need to continually re-establish an equilibrium by accommodating new elements within this environment: a kind of intellectual homeostasis. The view that the individual is capable of conceptual discovery through direct physical experience by manipulating material objects is an important justification for the use of individual laboratory work in science curricula. Another aspect of the nature of the learner is his readiness, reflected in Bruner's oft-asserted proposition that "it is possible to teach any subject to any child at any age in some form that is honest - and interesting" (1965, p.20). As observed by Shulman and Tamir (1973), this is a "peculiarly Brunerian version of Piagetian theory" (p.1099), but has nevertheless had important influences in both science and mathematics curricula.

A third theme discussed by Bruner is his 'learning-by-discovery' thesis which had been a central influence in the development of science curricula. The learning-by-discovery thesis has been subject to intense critical review and the claims for discovery learning have been investigated in many contexts. This research is discussed in a later chapter. Shulman and Tamir (1973) observe that although "it has repeatedly been noted that no firm evidence in support of the superiority of discovery learning exists" (p.1118), there are studies which suggest, and advocates, notably Schwab, who maintain, that where transferable problem-solving skills and attitudes are concerned discovery learning may still be a fruitful hypothesis.

One curriculum project of the time which faithfully reflected these major considerations of the Woods Hole Conference was the Biological Sciences Curriculum Study. In 1959 the American Institute of Biological Sciences, having secured substantial financial support from the National Science Foundation, set up the Biological Sciences Curriculum Study, and a professional staff was gathered under its director, Arnold B. Grobman, at the University of Colorado at Boulder to formulate basic policy. A Steering Committee was established which comprised college and university biologists known widely for their researches as well as their interest in educational problems, high-school teachers of biology, writers of textbooks and teachers' handbooks, and other educators. Consensus was achieved on a number of important issues which determined the character of the curriculum materials produced. Firstly, agreement was reached on the basic biological themes and concepts which should form the substantive framework of a high school course in biology (similar to Bonner's (1963) seven basic "ideas of biology"), but it was agreed that there was no one best way of organizing this substantive content. Shulman and Tamir (1973) observe:

> An important aspect of Schwab's work is his emphasis upon the substantive structure of a discipline as an imposed construction, not an organization inhering in that subject matter. It is no coincidence that the Schwab-influenced Biological Sciences Curriculum Study (BSCS) alone among the new curricula was produced in three versions, each reflecting a different way of conceiving the substantive structure of biology. (p.1102)

Secondly, it was agreed that science should be presented to students not as a fixed body of established laws and facts but as the fruits of limited observations and summations of experience. Science was conceived as "a body of methods of inquiry by means of accurate and confirmable observations, quantitative and mathematical analysis, and controlled experimentation" (Glass, 1964, p.96) and it was held that these should be the organizational principles of laboratory work. Thirdly, it was agreed

> that the boy or girl in school cannot comprehend the nature of science by learning facts about nature. Instead, real participation in scientific inquiry, and as full a participation as possible, should be provided. Only by engaging in the steps of scientific inquiry may a student become able to discern the true difference between sound experiment that provides evidence and complex instrumentation that offers a show - between evidence and authority, between science and magic. (Glass, 1964, p.98)

This distinction between inquiry-centredness and authority-centredness is a recurrent theme in the BSCS materials and is taken up in some detail in a joint publication by the Mid-continent Regional Educational Laboratory (McREL) and the BSCS, Inquiry Objectives in the Teaching of Biology (Bingman, 1969).

The following account of the introduction of the Web of Life course into Australian schools is given in Biological Science: Web of Life. Teacher's Guide (Australian Academy of Science [AAS], 1969):

> During 1964 teachers and biologists in Victoria and South Australia became interested in the American materials and as a result of preliminary trials it was decided to adapt them for use in Australian schools. The adaptation project was undertaken by the Australian Academy of Science with the cooperation of the Victorian and South Australian Education Departments. A team of five editors was appointed and the project approved by the BSCS. The designers of the Australian course have been able to draw on all the American materials and on the knowledge and experience of many biologists and teachers throughout Australia. (p.5)

The Web of Life course with its textbook, teacher's guides, and student's manuals is the result. It is interesting to note that the Australian course designers based the Web of Life course largely on the BSCS Green Version, which had an ecological emphasis considered to be appropriate for secondary school students. This outlook is shared by Ausubel (1966) who was very critical of the inordinate emphasis on experimental foundations and "an absurd level of biochemical and physiological sophistication" (p.177) found in the BSCS Yellow and Blue Versions. The adaptation is basically in terms of substantive content, that is, the Australian biome is the substance of the course. The commitment to inquiry-centred teaching strategies has been faithfully preserved. The following description of BSCS could equally well be applied to Web of Life. "BSCS Biology is not only a course of biology, but also a way of teaching biology with inquiry and discussion techniques essential strategies that were (sic) expected to be used by teachers" (Ramsey, 1975, p.107).

The introduction of the Web of Life course was based on the belief that the inquiry-based strategy gave a truer picture of the nature of biology and so should lead to a greater understanding of the nature of science, an increase in inquiry skills, and desirable attitudes towards science. Having different objectives from the pre-existing biology course, a different form of examination was required to test those objectives. Great care was taken by the examiners to develop a content-process grid to act as a table of specifications for the construction of the test; a much greater emphasis was given to items requiring higher-order cognitive skills and a reduced emphasis was placed on knowledge; and many test items were set in the context of biological investigations. In short, consistent with the expressed intentions of the course, the processes of inquiry were tested as well as the end products of inquiry.

In line with common social stereotypes of the role of women in society, during the first half of this century biology was regarded as a highly suitable science study for girls. It was assumed to be relevant to the nuturant role of women and a suitable preparation for paramedical occupations such as nursing, occupational therapy, and physiotherapy. In particular, it was assumed not to be as rigorous nor as intellectually demanding as the other commonly taught sciences, physics and chemistry, and it was accepted folklore that students who were also doing physics and chemistry would do very well in biology. However, after the introduction of the Web of Life course on a state-wide basis in 1967, teachers expressed concern to the examiners that many students who had performed well in the physics and chemistry examinations performed poorly in the biology examination. Implicit in most of the literature on discovery learning is the assumption that an inquiry-based strategy should also lead to greater achievement, so the observation that some students who have demonstrated high achievement in other subjects do not perform as well as expected in the Web of Life course was disquieting. The studies on inquiry-based strategies discussed in Chapter 2 present a confused picture, and numerous reviews suggest that the published research does not support the various claims made about the virtues of inquiry or discovery learning.

It appeared necessary to ask then, "Is Web of Life succeeding?" This was the question that prompted the present investigation although it was not quite the question that was finally asked.

One way of attempting to find out whether a particular instructional treatment is succeeding is to compare the results of that treatment with the results of the current, common, or traditional program. The conventional design is to assign students, if possible at random, to the alternative treatments and at the end of the program to compare average scores on some criterion test, often with covariance adjustment for prior knowledge or skills, and to find either that there are not significant differences or that one treatment is more effective than the other. But a consideration of the main effects alone tends to obscure the probability that there are some students who perform worse even under a treatment that is found to be 'best on average'. Some time ago Crandall (1963) wrote, "it has been reported by several other educational researchers that the personality predispositions of students may 'filter' the effect of various teaching innovations" (p.434), and concluded that there is likely to be an interactive effect between methods such as 'lecture discussion' and 'inquiry-training' and children's personalities on motivation and on performance. Cronbach (1966) commented:

> We should expect an interaction between the discovery variable and pupil characteristics .... I am tempted by the notion that pupils who are negativistic may blossom under discovery training, whereas pupils who are anxiously dependent may be paralyzed by demands for self-reliance. (p.90)

In major reviews of attribute-treatment interaction (ATI) discussed in Chapter 3, opinion ranges from unrelieved pessimism about the prospects of discovering significant interactions, through various degrees of caution to unbounded optimism. Nevertheless, it can be seen that there are constantly recurring suggestions by reviewers, and some experimental support for the hypothesis, that student characteristics such as general ability, prior achievement, personality attributes, and cognitive style interact with instructional treatments such as the degree of instructional support and other elements of teaching style.

There may then be some virtue in carrying out a study in which interaction effects are treated as the central concern and not as part of the error variance limiting the generalizability of main effects. The question to be answered then becomes "For whom is the Web of Life course succeeding and in the hands of which teachers?"

At the present stage, the research hypothesis can be stated in a general form: an outcome, for an individual student, is a function of the instructional treatment, the student's personological attributes, and the interaction between treatment and attributes. This hypothesis is elaborated after consideration of research relating to the variables, and the literature on experimental design, and is tested by multiple regression analysis as described in Chapter 8.

The dependent variables in the study are measured by two tests of biology achievement constructed for the Web of Life course by the Australian Council for Educational Research (ACER), and by four scales contained in a Biology Attitude Inventory (BAI) (Lucas & Broadhurst, 1972). A discussion of these instruments is to be found in Chapter 7.

With regard to the treatment variables, Ramsey and Howe (1969a) suggest that a far more important factor than contrasting teaching strategies may be the way in which the teacher works within the restraints imposed. The content of the Web of Life course, its objectives, and the recommended teaching strategies are described in detail in the curriculum materials designed for the course. However closely these teaching strategies are followed, individual teachers will place their own idiosyncratic imprint on classroom interaction (Yager, 1968), and it is the contrasts of natural teaching styles that form the alternative treatments in this study. The selection of contrasting groups of teachers and the validation of the differences in their teaching styles is described in detail in Chapter 6. Many of the reviews of ATI studies discussed later are critical of the brevity and artificiality of the instructional treatments. Snow (1974) observes that "most generalizations about school learning need to be built on research using substantial samples of learning time" (p.281). In this study the treatments consist of a full year's instructional program in biology by two groups of teachers with contrasting natural teaching styles.

The selection of student characteristics samples the major classes of personological differences described by Tyler (1969): achievement, which in this study is the dependent variable; intelligence and aptitudes; motivation and personality; and cognitive styles. The ACER Advanced Test B40 was used as a measure of general ability; a number of personality dimensions were derived from the Omnibus Personality Inventory (OPI), Form F (Heist & Yonge, 1962); and the Test on Cognitive Preferences (Mackay, Note 6) provided measures of cognitive styles. The discussion of the selection and measurement of these personological variables is to be found in Chapter 5.

CHAPTER 2

TEACHING THE PROCESS OF INQUIRY

2.1 BIOLOGY AS A PROCESS OF INQUIRY INTO THE LIVING WORLD

2.2 'DISCOVERY' AND 'INQUIRY'

2.3 TEACHING STRATEGIES USING THE LABORATORY

2.4 TEACHING STRATEGIES DESCRIBED AS INDUCTIVE OR DEDUCTIVE

2.5 TEACHING STRATEGIES DESCRIBED AS GUIDED DISCOVERY

2.6 CONCLUDING OBSERVATIONS

2.1 BIOLOGY AS A PROCESS OF INQUIRY INTO THE LIVING WORLD

The objectives of the Web of Life biology course have been largely derived from BSCS publications and show a strong emphasis on 'process' goals stemming from a basic view of "biology as a process of inquiry into the living world" (Teacher's Guide, AAS, 1969, p.6). This is not to say that the findings of inquiry (content) are regarded as unimportant, for as Bingman (1969) observes, "teaching inquiry processes demands the teaching of content inseparable from the process" (p.1). The development of inquiry skills is explicit in the statement of course objectives in the Teacher's Guide (AAS, 1973):

> Students should know how such ideas [the major ideas of biological science] are derived, preferably by experiencing scientific inquiry themselves, be familiar with evidence from which the ideas are derived, appreciate their implication, and be able to apply them to unfamiliar areas of biology, and to everyday problems. (p.7)
>
> Major ideas should emerge as a result of classroom activities which make clear the nature of scientific inquiry and teach the student to operate in scientific ways himself. Personal experience of scientific investigation is a very effective and desirable way of conveying the spirit and methods of scientific investigation. (p.7)

and in the statement of teaching strategies (AAS, 1973):

> The resource materials provided and learning situations employed should ... represent science as a process of inquiry (rather than as a rhetoric of conclusions), and use inquiry-centred teaching strategies where possible. (p.8)
>
> There is strong evidence that the objectives of the course can be best achieved by a course based on laboratory work which involves direct, personal observation and experimentation by the students. It is only by becoming active participants in the processes of science that students can fully appreciate what it is all about. (p.8)

The development of inquiry skills is also implicit in the teaching strategies which are known as Invitations to Inquiry.

> These are structured discussions in which the students are presented with data piece by piece and asked questions at each stage. Most Invitations serve a double purpose: they illustrate some aspects of the methodology of science, and they allow students to use or develop biological ideas. (AAS, 1973, p.15)

Teaching strategies appropriate to the aims of the Web of Life are discussed in some detail in Bingman (1969, pp.32-33, 44-84), Klinckmann (1970, chaps. 5-10) and in the Teacher's Guide (AAS, 1973, pp.6-23). In these discussions reference is made to research on teaching and learning in science education; in particular to claims made for individual laboratory work, for guided discovery, and for inductive rather than deductive methods in developing inquiry skills in science.

2.2 'DISCOVERY' AND 'INQUIRY'

As Shulman and Tamir (1973) observe, there is some ambiguity in the way in which the terms 'discovery' and 'inquiry' are used. In one usage the words are used as adjectives to describe independent variables such as methods of teaching, and are often contrasted with methods described as 'traditional', 'didactic', 'teacher-centred', 'dogmatic', or 'deductive'. In a second usage the terms refer to intervening variables, the processes which occur in the minds of students. In a third usage they are used to describe dependent variables or outcomes as in the initial paragraph in this chapter which refers to process goals and inquiry processes.

The description of methods as discovery or inquiry methods invites the reader to assume that these methods lead to the acquisition of the skills of inquiry or discovery, but obviously this relationship is one which has to be demonstrated. Also, methods which are described as inquiry methods may not live up to their description. Schwab (1962), Gagné (1963), Rutherford (1964)

Robinson (1965), Wittrock (1966), Bingman (1969), and Herron (1971) have all drawn attention to the methods of teaching intended to produce the skills of inquiry. Schwab, in particular, has articulated the distinction between science as inquiry and teaching as inquiry, and points out that the science classroom should be inquiring into the materials of science, which would thus demonstrate the process of inquiry.

Within the last decade there has been a number of comprehensive reviews of research in science education (e.g., Atkin & Burnett, 1969; Belanger, 1969; Ramsey & Howe, 1969a, 1969b; Novak, Ring & Tamir, 1971; Shulman & Tamir, 1973). Each review has had its own particular perspective, and each has cited studies which have touched on the process objectives or methods of instruction designed to achieve these. In the sections which follow it is not intended to carry out a similar comprehensive review, but merely to refer to individual investigations which illustrate the argument being advanced. The essential design of the studies is to evaluate the effects of contrasting instructional treatments using various criterion measures. The terms 'discovery', 'inductive', 'inquiry', and 'laboratory' typically refer to teaching strategies, and the outcomes are variously measured in terms of scores on tests of 'critical thinking', on 'understanding science', on attitudes to science, and on various achievement tests. The studies are grouped under the headings laboratory, inductive, and discovery as described by their authors.

2.3 TEACHING STRATEGIES USING THE LABORATORY

Many claims have been made for laboratory work as a means of achieving the aims of science teaching, and in particular for the development of inquiry skills (Brandwein, Watson & Blackwood, 1958; Pella, 1961; Teacher's Guide, AAS, 1973).

Atkin and Burnett (1969) refer to the very large number of studies which attempt to compare the results of teaching by lecture demonstrations with those of individual laboratory work. Comparisons of individual laboratory work and class demonstration methods have yielded few differences in achievement except on measures which require direct knowledge of or the use of laboratory procedures. It became clear, even in the early studies of Horton (1952) and Kruglak (1953, 1954, 1955), that students who have worked with laboratory materials develop greater skills in laboratory techniques and procedures than those who had not. However, these student behaviours are seldom sampled in typical tests of achievement in science, and paper and pencil tests have, not surprisingly, consistently failed to show significant differences in achievement in science between students who have been taught without practical work and those who have used the laboratory.

More recent studies are no more encouraging. Using two contrasting treatments to teach a unit on genetics, an 'inductive laboratory approach' and a 'traditional lecture-illustrative approach', Schefler (1965) reports no statistically significant difference in students' gain in knowledge of facts and principles, understanding of science, or attitudes towards science.

As the BSCS curriculum materials were designed to be investigatory in nature and to foster the spirit of inquiry, it is chastening to read that Yager, Engen, and Snider (1969), found that between 'discussion only', 'discussion-demonstration', and 'laboratory-discussion' methods of using the BSCS Blue Version materials, there were no significant differences in biology achievement, 'critical thinking', 'understanding science', or attitudes towards biology. Again, they found that students who worked with laboratory materials developed greater skills in laboratory procedures than those who had not.

Apart from the well-established and not-surprising finding that students who engage in laboratory work develop skills in laboratory procedures, laboratory work appears to achieve little more than lecture-demonstration methods. In particular, the weight of evidence suggests that too much may be claimed for the role of laboratory work in the acquisition of inquiry skills in science. However, the studies discussed do not establish beyond question that the laboratory was used to give students experience of "direct, personal observation and experimentation" (Teacher's Guide, AAS, 1973, p.8). While it may be that some teachers guide the students' own laboratory investigations, there may be many using these materials who still act as dispensers of knowledge, and use the laboratory only for drill and verification. An analysis of the laboratory manuals of the PSSC and BSCS courses carried out by Herron (1971) demonstrates a startling difference between the expressed intentions of these courses and the manner in which students are expected to work in the laboratory. In most exercises the problem was given, the ways and means of solving the problem were given, and the answers were given. However, Murphy (1968) discovered no significant difference between the content-centred and process-centred laboratories with respect to the development of knowledge, scientific attitudes, problem solving ability, or interest in biology. Similarly, Coulter (1966) found no differences in student achievement between inductive and deductive laboratory methods in biology for the learning of facts, application of principles, or for the utilization of laboratory techniques.

Without citing the evidence he claims that there was some indication that the inductive approach was, however, more effective in teaching cause and effect relationships, the making of judgements on evidence, and the evaluation of arguments.

Nevertheless, doubts such as those raised by Herron (1971), about the validity of describing some laboratory teaching as 'inductive' and other teaching as 'deductive', prompted Egelston (1973) to use interaction analysis to verify that the laboratory supervision was truly different in the two treatments. She found that "the use of an open-inductive method yielded significantly different results for laboratory behavior, for learning environment in the classroom and for academic achievement from [and superior to] that of the traditional deductive method" (p.476).

As well as the need for detailed description of the methods used in laboratory work, there may be considerable room for improvement in the criterion tests employed for measuring its outcomes. Tamir and Glassman (1970, 1971a, 1971b), setting their tests in the subject-specific context of school biology, have devised laboratory tests for BSCS students which investigate a much wider range of inquiry skills than measured by tests of manipulative skills or the Watson-Glaser Critical Thinking Appraisal. As Glaser (1966) himself observes "discovery requires different concepts and logical combinations when one is working in microbiology or botany, or breaking hieroglyphic codes" (p.23).

2.4 TEACHING STRATEGIES DESCRIBED AS INDUCTIVE OR DEDUCTIVE

From literature on inductive and deductive teaching strategies over 50 years, Dubois (1966) selects 12 investigations for critical review. Only 6 of the 12 investigations yield significant differences between induction and deduction as treatments, and learning, retention, or transfer as criteria. Of the 12 statistically significant measurements resulting from these 6 studies, 11 favour deduction over induction.

Krumboltz and Yabroff (1965), introducing their own study of a comparison of the effects of inductive and deductive methods of instruction, also refer to a number of studies with similar criteria of student achievement, some indicating that the inductive method is superior, some indicating that the deductive method is superior, and others suggesting that both methods are equally effective. In an attempt to resolve the problem, Krumboltz and Yabroff controlled two additional variables: the amount of participation by the learner, and the amount of information given to the learner prior to and following problem-solving activities. They concluded that the two methods were about equally efficient in the time taken to teach a given body of information, that the two methods were about equally effective in producing accurate transfer, and that although the inductive method produced slightly quicker learning of the "knowledge of rules" this advantage was offset by the dissatisfaction with that method by the groups using it. One significant interaction between verbal ability and treatment was reported.

It is apparent that even when studies comparing inductive and deductive teaching strategies look distinctly unpromising there are investigators who feel that the issue remains one of importance. The ensuing attempts at resolution often show more care in the experimental design, and particularly in the control of extraneous independent variables. As Coombs (1966) suggests, the inconclusive nature of the findings may be partly due to the fact that inductive and deductive methods as usually described are relevant only to the teaching of concepts and generalizations whereas many other items of content are taught in the classroom, and partly to the fact that many different types of behaviour are shown in classrooms which are not represented in descriptions of inductive and deductive methods. So it is possible that two instances of inductive teaching may differ more widely as instructional methods than an inductive method differs from a deductive method.

Different methods are difficult to describe with precision and often cannot be easily dissociated from the particular subject matter being taught, or from the teacher's personality. The investigation of Sakmyser, (1974) would appear to have overcome these difficulties. Two versions of programmed instruction were devised for teaching the concept of chemical equilibrium. Even here "neither the inductive nor the deductive approach ... was significantly more successful in teaching the concept .... This finding is consistent with earlier research" (p.75). An interesting outcome of Sakmyser's study with particular relevance to the present investigation is that each program was shown to be superior for individuals with certain personality traits or skills. Interaction effects were found with each of the second-stratum factors of Cattell's High School Personality Questionnaire.

In their review of trait-treatment interaction studies which used inductive-deductive strategies of teaching as alternative treatments, Berliner and Cahen (1973) observe that

> it is not too early to note that generalizations about the utility of inductive or deductive teaching methods are likely to be oversimplified. This oversimplification will continue until the educational research community accepts the caution that there is a high probability that students with different traits are likely to show differential achievement under one or the other approach. (p.72)

They note, for instance, that although Anderson's (1967) review clearly points out the general superiority of rule-example methods over example-rule methods, for some sub-groups of students this generalization does not hold.

2.5 TEACHING STRATEGIES DESCRIBED AS GUIDED DISCOVERY

An example of a study which has concentrated on the acquisition of concepts and generalizations is that of Boleratz (1967). Her experimental instructional treatment consisted of a program designed to emphasize discovery and the transfer of generalizations, and the control treatment consisted of teacher-designed conventional learning experiences. She found that students' performances on her Value Concepts Test were independent of IQ, socioeconomic status, or the experience of their teacher, but that those who participated in the experimental learning program performed better than those in the control group. However, the amount of guidance given to the 'discovery-method' and 'control' group teachers differed greatly. To help set a standard for the discovery method treatment, teachers were provided, for guidance, with detailed lesson plans containing a list of objectives, illustrations, case studies, and the like. Although the lesson plans of the control teachers were checked by the experimenter they were not given the assistance given to the discovery method group. This study is in marked contrast to that of Ripple, Millman, and Glock (1969) where the alternative treatments were each made as good as was possible.

Thomas and Snider (1969) compared the effects of 'guided discovery' versus 'didactic' teaching methods on the acquisition of inquiry skills. They found that the didactic method was to be favoured for factual-conceptual outcomes as measured by a locally designed test and the TAB Inventory of Science Processes, and the guided discovery method was to be favoured for inquiry skills as measured by the Watson-Glaser Critical Thinking Appraisal. They also found significant interaction of treatment with intelligence in the development of inquiry skills: guided discovery was better for the high intelligence group and the didactic method was better for the low intelligence group.

These two studies illustrate the divergent nature of findings characteristic of investigations into discovery learning, and Thomas and Snider's study illustrates the importance of congruence between the criterion measures and the purposes of the treatment. Is it any wonder that didactic teaching, if true to its description, is better than guided discovery in leading to factual-conceptual outcomes, or that guided discovery, if true to its description, is better than didactic teaching for skills measured by the critical thinking test? The different methods are based on different assumptions about what is important in learning: the didactic method traditionally places high importance on the acquisition of knowledge, and discovery learning on the acquisition of inquiry skills.

The extensive literature on discovery learning has been subject to critical review by the contributors to Shulman and Keislar (1966), and also by Belanger (1969), Hermann (1969), and Strike (1975). Wittrock (1966) asserts that "the aging but still elusive learning-by-discovery hypothesis has outlived its usefulness" (p.74). Strike (1975) agrees that research on discovery learning has been inconclusive, and believes that "empirical studies, even conducted with unimpeachable rigor, have failed to agree because they have not shared a common interpretation of either dependent or independent variables" (pp.461-462). Belanger (1969) also comments on the wide range of meanings assigned to the words 'discovery' and 'inquiry' which makes generalizations hazardous, but concludes that the published research leaves unsubstantiated the various claims made for inquiry or discovery learning. Nuthall (1973) concludes that the model-dominated research which pits a 'new method' against the 'traditional method' (which is always assumed to be bad) is unlikely to succeed because it tends to make a priori claims about what teaching is like, and because, without careful examination, it elevates the 'method' to the position of a "valid empirical entity" (p.36).

2.6 CONCLUDING OBSERVATIONS

In summary, research on the efficacy of contrasting teaching strategies in leading to either process or product objectives presents a confused picture.

There are very few replicated investigations. Experimental and field studies of teaching science in the secondary school are usually individual studies and seldom those of research teams or institutions, and so it is to be expected that each study tends to break new ground. Different criterion measures are used, and many designs control for different combinations of extraneous independent variables. Different combinations of treatments are compared and as suggested by Coombs (1966) there is little assurance that methods described in similar terms, such as 'traditional lecture-discussion' or 'laboratory-demonstration', are in fact, strictly comparable from study to study.

It is clearly important to establish what the basic assumptions underlying and the purposes of a particular instructional strategy are, and not to attempt to compare strategies which have different aims, either by using criterion tests which are relevant to neither set of aims or by using tests which favour one set of aims rather than the other. On the other hand it may be said that, if a course has an agreed set of aims, it is legitimate to compare the effectiveness of alternative strategies for achieving these aims, but it is not a fair race if only one of the horses, or neither, is trying to win.

Another problem with most of the studies quoted above is the lack of an adequate description of the treatment variables: a description which substantiates that the laboratory is used in an investigatory manner, or that a so-called inductive method is true to its description and differs from a deductive method. A related difficulty is in knowing precisely what are the contrasting elements in the alternative treatments: are they elements in the teaching strategies, are they elements in the subject-matter content, or are they elements in the teacher's personality or teaching style? Ramsey and Howe (1969) suggest that a far more important factor than contrasting teaching strategies may be "the philosophical approach of the teacher and the way in which he works within the restraints imposed" (p.63). This suggestion receives support from a study of Boeck (1956). In investigating three different instructional treatments, he found that individual teachers exercised greater influence over pupil achievement than did the various treatments.

Both Wittrock (1966) and Cronbach (1966) draw attention to two important aspects of most investigations of contrasting instructional treatments. Firstly, many studies are of comparatively short duration and there is considerable suspicion that the outcome may have been different if the treatment had been of longer duration. Secondly, any particular teaching strategy may be expected to be better for some students than for others. Bennett (1976), in referring to a number of ATI studies, says:

> Such interactions are thought to be one reason why comparisons between teaching methods in the past have produced results indicating no difference. 'It is not that the different methods are interchangeable - rather that each is successful and unsuccessful with different kinds of pupils' (Leith, 1972). (p.25)

Cronbach (1966) asserts that he has

> no faith in any generalization upholding one teaching technique against another, whether that preferred method be audiovisual aids, programmed instruction, learning by doing, inductive teaching or whatever. A particular educational tactic is part of an instrumental system; a proper educational design calls upon that tactic at a certain point in the sequence, for a certain period of time, following and preceding certain other tactics. No conclusion can be drawn about the tactic considered by itself. (p.77)

In the studies quoted above, which were designed primarily to test for the main effects of the treatment variables, there are occasional suggestions of interactions between instructional treatment and pupil characteristics (e.g., Thomas & Snider, 1969; Sakmyser, 1974). Other studies which point to attribute-treatment interaction are discussed in the following chapter.

CHAPTER 3

ATTRIBUTE-TREATMENT INTERACTION

3.1 INTRODUCTION

3.2 GENERAL ABILITY AND INSTRUCTIONAL TREATMENT

3.3 SPECIAL ABILITIES AND INSTRUCTIONAL TREATMENT

3.4 PERSONALITY ATTRIBUTES AND INSTRUCTIONAL TREATMENT

3.5 COGNITIVE STYLE AND INSTRUCTIONAL ACHIEVEMENT

3.6 PRIOR ACHIEVEMENT AND INSTRUCTIONAL TREATMENT

3.7 CONCLUSION

3.1 INTRODUCTION

Interactions between treatment variables and student aptitudes, attributes, traits, or characteristics, have been reported in the research on contrasting instructional treatments, but until the last decade these interactions were regarded almost as part of the error variance rather than as a basic discovery to be interpreted.

The investigation of aptitude-treatment interaction (ATI) as a research strategy in education was suggested by Cronbach (1957), when he proposed that the interests and approaches of the correlational psychologists concerned with individual differences and those of the experimentalists concerned with the search for general laws of learning should be brought together. An exclusive emphasis on the search for the 'one best treatment' applicable to 'persons-in-general' would appear to ignore the realities of the rich complexities of human behaviour. The personological variables which characterize each individual result from one assortment from the limitless possibilities of genetic recombination, which, in turn, enters into a unique interaction with a complex environment. Although the roles of the teacher and the student are, for the most part, clearly established, and although the processes of socialization succeed in producing a measure of conformity, each individual will fulfil his own role in a unique manner. Gagné's (1967) symposium on learning and individual differences was an important statement reaffirming the concern of educators with adaptation to individual differences: a number of contributors (Anderson, Cronbach, Gagné, Glaser, Jensen) suggested that the same instructional treatment can not be expected to prove best for all students.

Newsom, Eischens, and Looft (1972) argue, as did Cronbach (1957), that there is a great need to reconcile the nomothetic and idiographic approaches to the study of behaviour. They comment on the pre-occupation shown in experimental research with the search for general laws of learning, and refer to the different conclusions reached about incremental learning from the study of groups and from the study of individuals. They observe that

> differences among individuals in learning are ... not to be conceived as solely due to dissimilarities in some unidimensional quality which reflects the efficiency of the processing mechanisms, but rather they are to be considered as the result of various conditions of the learning situation interacting with idiosyncratic structures and mechanisms of the learner. (p.387)

They conclude that an individualized program for every student would be the superior method, but that full implementation of an entirely individualized instructional program would be an unrealistic hope. They suggest that a plausible approach would be to develop programs for 'prototypic learners' so that

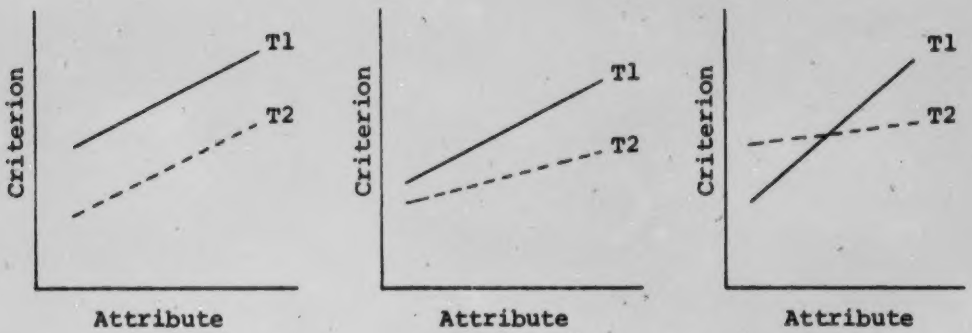
> students with common interests and similar individual difference profiles, could then be provided with materials according to whatever constellations of these variables are deemed important for mastering that given content area or skill. (p.391)

The two most recent reviews on ATI are those by Hunt (1975) and Tobias (1976). Tobias refers to the adoption, by various authors, of different terminology to describe this research: aptitude-treatment interaction, trait-treatment interaction, person-environment interaction, and attribute-treatment interaction. He argues that the term aptitude has

> so traditionally been associated with the cognitive domain that it may well predispose investigators to concern themselves primarily with variables such as scholastic aptitude or intelligence, (p.62)

and that the term trait "suggests a stable predisposition, not infrequently in the personality domain, relatively unchanging over short periods of time" (p.62). Both Tobias and Hunt favour Rhetts' (1972) term of attribute-treatment interaction in that it retains the widely used abbreviation ATI and does not suggest a limitation to the range of variables studied. Throughout this study the preference of Rhetts, Tobias, and Hunt will be followed for the reasons they have expressed.

Cronbach and Gleser (1965) have pointed out that information on differential student attributes is of use in providing differing treatments only when the attribute and treatment can be shown to interact. Some possible relationships between a particular student attribute and the effects of two instructional treatments as measured on some criterion test are illustrated in Figure 3.1.1.



3.1.1a 3.1.1b 3.1.1c

Figure 3.1.1. Relationships between student attributes and instructional treatments.

In Figure 3.1.1a, measures of the student attribute correlate positively with success for both treatment groups, and the regression lines indicate that treatment 1 is superior to treatment 2 but that there is no interaction between the student attribute and treatment. Similarly, in Figure 3.1.1b the attribute correlates positively with success for both treatments, but the regression lines have different slopes, although one remains above the other throughout the range of the attribute measured. Such an interaction is termed ordinal and is distinguished from a disordinal interaction (Lubin, 1961) shown in Figure 3.1.1c, where the regression lines for treatments intersect within the range of the attribute measured.

The conventional wisdom is that the ordinal interactions give no basis for assigning students to the alternative treatments because one treatment is best for all classes of students (Ripple, Millman, & Glock, 1969, p.119), and that where disordinal interactions occur, students to the left of the intersection should be assigned to one treatment and students to the right of the intersection should be assigned to the other (Cronbach & Snow, 1969, p.18). As Cronbach and Snow (1969) observe, ordinal interactions "can be eliminated if either variable is subjected to some particular monotonic transformation" (p.20). Where scales are arbitrary, as in most educational research, it makes sense to ignore effects that arise from the scale. However, they give two reasons for treating ordinal interactions more seriously. Firstly, where the outcome scale can be transformed by a decision-maker into non-arbitrary terms such as cost, or availability of resources, one treatment may only be available to a limited number of students. In such a case the interaction can support "a decision rule to assign to the more costly treatment only those persons for whom there is clear chance of benefit" (p.19). Secondly, where ordinal interactions are established the regression lines may be expected to intersect in regions above or below the levels of the attribute measured. Extrapolation of the regression function from one sub-population to another can provide useful hypotheses for testing.

Bracht and Glass (1968) and Bracht (1970) place an even more stringent restriction on disordinal interactions than does Lubin (1961). They claim that Lubin's distinction does not provide sufficient protection against a Type I error, and specify that an interaction effect should be "operationally defined as disordinal only when the differences between alternative treatments at two levels of a personological variable are both significantly non-zero, and different in sign" (Bracht, 1970, p.632). Berliner and Cahen (1973) feel that this additional requirement is "overly conservative and perhaps unnecessary" (p.61). Bracht (1970) critically examined 90 studies which were designed to permit a test of interaction between personological variables and alternative treatments, and found that of the 108 relationships examined in these studies only 5 indicated disordinal interactions. In Bracht's review the more stringent requirement of Bracht and Glass (1968) was used, and the other 103 relationships were grouped together as showing "ordinal or no interaction". Although this review gives little support for the interaction hypothesis it may be slightly less damaging than it appears at first. The quality of the studies included in the review varies widely, many not being designed to test for an attribute-treatment interaction. Also, there is no indication of how many studies showed significant ordinal interactions, nor how many studies demonstrated disordinal interactions that were judged by the reviewer to be non-significant when, under less conservative criteria, they may have been judged significant. Bracht acknowledges that the results of the review are "primarily descriptive and have limited inferential value" (p.635).

Hunt (1975) considers a number of reasons for what he regards as resistance to the adoption of an interactive paradigm by educational psychologists.

He claims that the core of the resistance is due to the search for general principles, and in this context interactions pose a threat to the validity of such principles. However, the results of an experiment are specific to the characteristics of the persons investigated and can be regarded as an example of a particular person-environment combination rather than as a base for wider generalization. The search for principles which apply to 'persons-in-general' produces other strange consequences: Hunt refers to the data on which Hovland (1937) based his principle of stimulus generalization. The mean scores on response amplitude decreased in an orderly fashion as the similarity from the initial conditioned stimulus was decreased. However, if individual scores are examined this principle is only true for one individual: the scores of only one out of 20 subjects showed a consistently decreasing pattern. The search for general principles has also led to the tendency to select personological variables which are consistent and stable; in other words, relatively unaffected by environmental variation. Glaser (1972) has suggested there is a strong case for educational researchers seeking to discover and refine 'new aptitudes' which describe students' differential response to variations in the instructional environment.

The second reason for resistance to the person-environment paradigm comes from the belief that not many ATIs exist. Hunt claims that ATI has become the domain of a special interest group, and that this has discouraged others from looking for differential effects. Furthermore, the stringent requirements of significant disordinal interaction proposed by Bracht (1970) have led most reviewers to be extremely cautious about claims that aptitude-treatment interactions have been demonstrated; cautious to the point where even their own credulity was stretched. Cronbach and Snow (1969) comment:

> One reaction ... would be to abandon ATI research on the grounds that such effects are non-existent. We urge against this defeatist course. It is inconceivable to us that humans, differing in as many ways as they do, do not differ with respect to the educational treatment that fits each one best. (p.193)

Glaser (1972) adds that the conclusion that few or no ATI effects have been solidly demonstrated is "astounding" (p.8). It appears that investigations conforming to the patterns and techniques conventionally described as ATI research have produced little evidence of interaction, but it is a fact of common experience that students do react differently to different ways of teaching.

Both Hunt (1975) and Cronbach and Snow (1969) refer to reservations held by some people about differentiation between students in the provision of alternative instructional treatments. Some fear that differentiation may be easily translated into value dimensions of 'good' and 'bad', and that as in the parable of the talents, "to everyone who has will more be given, but from him who has not even what he has will be taken away" (Luke, 19:26, RSV). Others fear that to assign students to programs they can handle may encourage the avoidance of a major educational responsibility, that of remedying deficiencies. These reservations, it must be noted, apply not to the principle of differentiated instructional procedures, but to its insensitive application.

In Lewinian (Lewin, 1935) terms, Hunt (1975) proposes a Behaviour-Person-Environment paradigm which he believes does not suffer from the restrictive influences described in previous paragraphs. The four characteristics of the paradigm are that:

> (1) it should be interactive not only in co-ordinating person-environment interaction, but in accommodating differential behavioral effects; (2) it should view the person in developmental perspective so that the differential effects of environmental influence may be seen developmentally as well as contemporaneously [a concern that has been voiced by Mitchell (1969, p.703),

> and Berliner and Cahen (1973, p.68)]; (3) it should consider person-environment interaction in reciprocal terms that view the effect of the person on the environment as well as the effect of the environment on the person; and (4) it should consider practical implications of such interactions so that the conceptions can be enriched by application. (p.218)

Using the personality construct 'conceptual level' (Harvey, Hunt, & Schroder, 1961; Schroder, 1971; Schroder, Driver, & Steufert, 1967) as the personological variable, and the degree of structure or organization of the learning environment as the treatment variable, Hunt develops a model based on the matching of conceptual level and degree of structuring. He illustrates the four characteristics of the paradigm by taking examples from the conceptual level matching model and suggests that the paradigm "permits conceptualizing ... otherwise impossibly complex problems in a form that may lead to understanding" (p.227).

Cronbach and Snow (1969) provide a comprehensive discussion of many aspects of ATI research and a critical review of a number of studies which bear on general ability and its possible interactions with treatment variables, specialized abilities and their possible interactions, and interactions in the personality domain. The three sections of their review are considered below, together with related recent studies.

3.2 GENERAL ABILITY AND INSTRUCTIONAL TREATMENT

Cronbach and Snow reviewed nine studies on programmed instruction where the alternative treatments required an overt response to the programmed mode, or a covert response (either reading from a text or from a program with the blanks filled in). Only one of these studies showed a significant interaction: Wittrock (1963), comparing posttests of those with high and low mental age, found those with low mental age were helped by having to respond overtly whereas those with high mental age were slightly handicapped. Overall however, the reviewers conclude that "as to interactions, the verdict is essentially negative" (p.85).

In addition, eight studies were reviewed where the alternative treatments were 'smooth' programs (those where the program is written in small steps and in carefully ordered sequence), or 'rough' programs (those where the step size was greater or the sequencing less orderly). It has been hypothesized that a high degree of structuring of programs would be advantageous to those persons with limited insight. Although two studies suggest that the rough programs are superior for abler pupils the interactions are either weak or non-significant. A large-scale study by Maier and Jacobs (1966) which took the class as the sampling unit found that the rough programs were superior for abler classes although when the data were re-analysed at the pupil level no interactions were found within the groups. Cronbach and Snow comment that this study gives a hint that "interaction effects may be mediated by the reaction of the class as a whole" (p.89).

Another study on programmed instruction by Maier and Jacobs (1964, cited in Cronbach & Snow, 1969) has some features of particular interest to the present investigation. The basic finding of the study was that students of high IQ do best with a combination of programmed instruction and a teacher who favours innovative methods, but that students of low IQ do best when taught by a conventional teacher without programmed instruction. Of more interest however, is the reviewers' evaluation of the study.

> The many excellent features of this study should be noted: reproducible treatments (at least insofar as the PI component was concerned); use of two treatment dimensions (extent and character of live teaching); use of more

> than one aptitude variable and use of more than one outcome variable; instruction extending over a realistically long time with real class material; and a large sample (77 classes with some 900 students). It will be noted that here again the investigators employed the class as the sampling unit, whereas nearly all educational investigators have employed persons as the unit. This is technically correct, but it means that the authors have labelled nonsignificant many effects that would have been called significant by most of those who report on these matters. Their nonsignificant effects are often large, and if confirmed with more classes could be important. No doubt a cumbersome multivariate analysis that examined both between-class and within-class regression would extract more information from a study with so complex a design; but the state of our insight into the interaction is not far enough developed to make good use of the statistics toward which the field will surely move. (Cronbach & Snow, 1969, p.96)

A review of three further unpublished studies (Stallings & Snow, n.d.; Salomon, 1968; and Koran, 1969), which showed interaction between general ability measures and alternative treatments, led Cronbach and Snow to suggest the following interpretation: treatments which prove beneficial to students deficient in some particular aptitude may function by providing the information-processing functions that the learner cannot provide for himself. In the three studies reviewed the treatments that were beneficial for the group of lower ability appeared to give emphasis to detailed drill that isolates stimulus elements and uses concrete representations, and may in fact be providing coding systems for the learner.

This interpretation receives further support from three recent experiments carried out by Salomon (1974). In dealing with the ability to visualize, Salomon found that students of low ability profited most from treatments giving high instructional support, but that students of high ability performed best under conditions of minimal support.

3.3 SPECIAL ABILITIES AND INSTRUCTIONAL TREATMENT

Differential abilities have long been used as a basis for educational guidance and may be expected to be an important source of interaction with instructional treatment. Cronbach and Snow reviewed 11 studies using specialized abilities as independent variables. Only 2 studies show interactions with differential aptitudes although another 4 show interactions of treatment with general ability. The reviewers comment:

> There are several interactions but it is not at all clear why they happened. There are no successes in planning treatments to capitalize on spatial or figural ability, though two studies with uncontrolled treatments reported interactions of outcome with these special aptitudes. (1969, p.141)

A recent study by Ott and Macklin (1975) analyses the achievement of first year college physics students in two instructional treatment groups: an 'audio-visual' treatment and a 'lecture-recitation - laboratory' treatment. They found indications of a strong disordinal interaction between the treatments and mathematical aptitude, as measured by the mathematics portion of the Scholastic Aptitude Test (SATM), and between the treatments and mathematical achievement, as measured by a pretest of the investigators' own design. They found that the standard treatment was more effective than the audio-visual treatment for students with mathematics pretest scores and SATM scores at the top end of the ranges, and that the audio-visual treatment was more effective than the standard treatment for students with mathematics pretest scores and SATM scores in the lower half of the ranges.

3.4 PERSONALITY ATTRIBUTES AND INSTRUCTIONAL TREATMENT

The diverse collection of studies reviewed by Cronbach and Snow involving personality attributes is organized by them into two broad groups. The first group of studies is concerned with tension and fearfulness and involve the broad complex of variables such as anxiety, neuroticism, compulsiveness, confidence and so on, and the second group is concerned with ambition, energy, and goal seeking to which Cronbach and Snow have given the blanket term 'constructive motivation'.

It is observed that the data from most studies in the first group are consistent with a curvilinear relationship between the attribute measured, say anxiety, and treatment, in that the outcome is poor at high and low levels of anxiety and is optimum at some intermediate level. Where alternative treatments are used, the optimum for one treatment comes at a different point on the anxiety scale than for the other. The treatment that is more effective at higher levels of anxiety generally contains less stressful elements. If some criterion performance has to be mastered, and an individual is not aroused then some 'stress-producing stimulation' may be beneficial. If the individual is aroused, then additional stimulation may be counter-productive. Considering the weight of evidence on curvilinear relationships between personality characteristics, such as those just mentioned, and test or instructional treatments, caution should be used in interpreting the results of studies which report only means for high and low groups. 'High' and 'low' groups may be drawn from extreme groups in the distribution or may be separated at the median of the sample, so that meanings differ between studies. 'High' and 'low' may have different consequences in different parts of the range and comparisons between the two may have limited meaning. Data interpreted as showing linear regressions may in fact be drawn from one limb of the arch-shaped curve, and data interpreted as an absence of relationship may be drawn from opposite limbs.

Among those studies in the first group which used the degree of structuring of instruction as treatment variables, are found repeated suggestions that students who are above the class median in anxiety respond better to an instructional treatment with more structure. The studies vary in their definition of the personality constructs investigated and in the way in which 'structure' is defined so that more assertive generalizations would be imprudent.

Many of the studies used extremely short instructional treatments, but that of Grimes and Allinsmith (1961) utilized ongoing school programs. In order to test the hypothesis that students' personalities affect their ability to profit from alternative methods of teaching reading, Grimes and Allinsmith went into two school systems, one where a phonics method was used and the other where a less structured method was in use. An attempt was made to select schools of similar socio-economic class background. Third grade students were tested for anxiety and compulsivity; these two measures were uncorrelated. Although the structured method produced the better overall result, there were significant interactions between personality and instructional treatments. The structured program was beneficial to those students who were anxious and/or compulsive; those who were less anxious and compulsive were more productive in learning situations where less structure was imposed.

Several studies were reviewed by Cronbach and Snow which investigated possible interactions of anxiety with task difficulty. The hypothesis under examination is that where the learning task requires only one response students with high drive should perform well, but where the task is more complex allowing for several competing responses a high level of drive may be disadvantageous. The studies reviewed give support to this hypothesis with the exception of those of Katahn (1966). The studies concerned easy and difficult serial learning tasks by college students. In one study there was no demonstrable relation between learning and anxiety and in the other study the high-anxiety students did best on the difficult task. The reviewers comment that "it is possible that Katahn got an unexpected result simply because his students did not find the task as difficult as he expected" (p.161).

The second group of studies concern attempts to identify interactions of constructive motivation such as need for achievement. Although the five studies reviewed indicate interactions, Cronbach and Snow are generally critical of their design or analysis. Domino (1968), by analysing the teachers' instructional procedures, classified 105 classes as either rewarding conforming behaviour, or rewarding independent behaviour on the part of students. Students were selected from the extremes of the distribution of two of the California Personality Inventory (CPI) subscales: Ac (achievement through conformity) and Ai (achievement through independence). Four groups of 22 subjects each were matched for sex and a non-verbal general ability test score. The data suggest that high Ai students do better in classes which reward independence and low Ai students do better in classes which reward conformity. There appears to be no interaction with Ac. The reviewers, although commenting on the weakness of the relationships demonstrated, suggest that the study should encourage more powerful studies of a similar nature. McKeachie, Lin, Milholland, and Isaacson (1966), investigating student affiliation motives, teacher warmth, and student achievement, conclude that ATI have been demonstrated, but Cronbach and Snow accept this conclusion cautiously in view of the "inconsistency of results and their generally small magnitude ... [and the absence of data on] the total number of relationships from among which the 'significant' results came" (1969, p.166).

Doty (1967) investigated the effectiveness of alternative instructional treatments in relation to five student characteristics: cumulative GPA, creativity, achievement need, social needs, and attitude towards the teaching method. She found that students who achieved most in the conventional lecture situation were characterised by moderate prior achievement and social needs and low creativity. Those who performed best with taped lecture conditions had low social needs and moderate achievement need and were low in creativity. She comments that the findings of this study and that of Doty and Doty (1964) suggest that for some students social interactions may be an important incentive to successful academic performance.

Snow, Tiffin, and Siebert (1965) also used prior achievement, personality, and attitude measures in investigating possible interactions with filmed versus live lecture demonstrations in college physics. Using measures of immediate recall as criteria they found that scores on two scales of the Gordon Personal Profile, Ascendancy and Responsibility, interacted with instructional treatment. The more submissive and more responsible students appeared to profit more from filmed demonstrations, the more assertive and less responsible students performed better with the live demonstrations.

A carefully designed study published after Cronbach and Snow's report is that of Ripple, Millman, and Glock (1969). The study, subtitled 'a search for disordinal interactions', utilized two widely different instructional modes (programmed instruction, and conventional instruction) and four student characteristics that were to be particularly congruent with the differential nature of the treatments (anxiety, compulsivity, exhibitionism, and convergent/ divergent thinking). In each of 22 schools pairs of English classes were roughly matched on the distribution of mental age and sex, and were randomly assigned to one of the instructional treatments. The programmed treatment comprised 10 lessons on vocabulary development. In the conventional treatment teachers were provided with 10 lesson plans designed to correspond with the content of the programmed lessons, student-teacher verbal interactions were heavily emphasized and a rich variety of supplementary verbal and audio-visual materials was provided. Both treatments were as good as they could be made. Great attention was paid to the preparation and refining of three factor-analytically derived criterion tests. Separate four-way analyses of variance were carried out for each student characteristic for each of the four criterion measures: three test scores and a total score. There was no significant interaction between learner characteristics and the type of learning task structure. In all cases the conventional-instruction group performed better than the programmed group. In view of the care taken in choosing student characteristics that were theoretically related to the treatments, in developing and carrying out the alternative instructional treatments, and in the design of valid criterion tests, this search for disordinal interactions has produced disappointing results. However, it is important to note that students spent about 15-20 minutes each day on programmed lessons, while the conventional group utilized the entire class period of 40-45 minutes each day. This additional time spent by the conventional group and the excellence of the teachers, or lesson plans, might all help account for the superiority of the conventional instruction group, which in turn contributes to the lack of disordinal interactions.

However, two recent investigations which have attempted to match features of the instructional treatments closely with the variations in student attributes have been more successful in finding interactions. Daniels and Stevens (1976) classified 68 students taking an eight-week introductory course in psychology as reacting to either an internal or external locus of control. Approximately half of each group were given a traditional teacher-controlled method of instruction with norm-referenced grading and the other half of each group entered into contracts, their grades being determined by the work (and make-up work without penalty) they were prepared to put in. All students were given a 75-item multiple-choice posttest. Analysis of covariance detected a strong disordinal interaction: internals performed better under the contract plan, and externals performed better under the teacher controlled method. Hunt (1975) proposed the matching of instruction with students' conceptual level: students with a low conceptual level (dependent on external standards and incapable of generating their own concepts) should profit more from instructional treatments with a high degree of structuring by the teacher, and students with a high conceptual level (capable of holding internal standards and of generating new concepts) should profit more from low structure, or be unaffected by structure. He found support for this hypothesis in a number of investigations, including a study of lecture (high structure) and discovery (low structure) teaching, and rule followed by example (high structure) and example followed by the rule (low structure).

Tallmadge and Shearer (1969) hypothesised that student characteristics might interact not only with instructional technique but also with the nature of the subject-matter content. Two contrasting subject-matter areas were chosen: Transportation Technique which is governed by logically-consistent inherently 'meaningful' rules, and Aircraft Recognition which is governed by 'arbitrary' rules. Two separate courses were designed for each of these subject-matter areas: one which was essentially inductive in nature, stressing discovery procedures and an understanding of concepts and principles, and the other was "deductive, didactic, and rote" (p.223). Each of these four courses was administered to 50-60 Navy men. As in Ripple, Millman, and Glock's study, measures were obtained on student characteristics that "were selected to represent those 'factors' which were judged to be closely related to the instructional methods selected for investigation or to the selected subject-matter areas" (p.223). Correlational techniques were used for 28 measures of aptitude, interest, and personality and criterion test scores within each treatment group. The study strongly supports the existence of learning styles, the correlates of which are non-cognitive rather than cognitive characteristics. The nature of the content was found to be a critical factor in affecting the magnitude and direction of the relationships existing between learner characteristics and instructional treatment. The authors conclude that students with the first learning style (characterised by high scores on the Scientific Interest scale from the Kuder Vocational Preference Record) showed improved learning of 'logically-structured' material taught inductively, and 'arbitrary classificatory' material taught deductively. Those with the second learning style (characterised by low scores on the Kuder Scientific Interest scale) showed improved learning under the reverse treatments. However, as the authors and Salomon (1972) point out, although a distinction was made between the alternative subject-matter as being characterised by 'meaningful' rules or 'arbitrary' rules, there may have been other differences that contributed to the effects that were found.

This interest in learning styles and their interaction with instructional treatments has been pursued by Cunningham (1975). He refers to "the inconsistent and general nature of much of the findings related to teacher effectiveness" (p.171), and observes that it may not be possible to discover factors common to all good teaching. He tested 1291 kindergarten students on a number of cognitive and personality characteristics and by a Q-mode factor analysis identified four types of students accounting for 94.4% of the variability in the measures used. Similarly, from data on teacher characteristics including beliefs towards instructional strategies, the 108 teachers were categorized into one of four types. His purpose was to determine if different student/teacher pairings resulted in significantly different academic performance for students. Student success was defined as the differences between observed scores on language, mathematics, and task orientation tests and predicted scores adjusted for age, sex, race, together with pretest differences in pre-school readiness, intelligence, home environment stimulation, language concepts, mathematical concepts, introversion/extroversion, social behaviour, and task orientation. Stepwise multiple-regression analysis was used first to identify the set of independent variables which were the most effective predictors of success. In an analysis of variance no significant differences were demonstrated between teacher types, but differences between student types and the pairing of student type with teacher type were found to be significant. His general finding was that "certain types of teachers are significantly more effective with one type of student than another and that all student types seem to have at least one type of teacher with whom they achieve significantly greater success" (p.186).

3.5 COGNITIVE STYLE AND INSTRUCTIONAL TREATMENT

Tallmadge and Shearer's (1969) use of the term 'learning style', Cunningham's (1975) clustering of students into four types, and Hunt's (1975) investigation of students' conceptual levels appear to herald a new move to characterize students in terms of their ways of learning. The concept of 'cognitive style' is treated in more detail in Chapter 5 but it may be useful to consider, in the context of ATI research, some studies using various measures of cognitive style.

Glaser (1972) points out that the aptitude-treatment interaction studies reviewed by Bracht (1970), Bracht and Glass (1968), and Cronbach and Snow (1969) had not convincingly demonstrated many interactions between treatment and the traditional measures of individual differences in aptitude, and that the "generally used aptitude constructs are not productive dimensions for measuring those individual differences that interact with different ways of learning" (p.8). Glaser argues that intelligence and aptitude factors have characteristically been selected because of their known correlations with instructional outcomes and not because of their relationship to learning processes. He points out that a variety of cognitive processes can be identified experimentally and suggests that these 'new aptitudes' may be far more productive variables. As support for the study of process concepts as individual difference variables he cites work on the reflection-impulsivity dimension by Yando and Kagan (1968), and Meichenbaum and Goodman (1969) and on cognitive styles by Kagan and Kogan (1970). Di Vesta (1973) also argues that, in the investigation of trait-treatment interactions, students' preferences for employing a learning strategy should be considered along with their abilities to apply the strategy.

Siegel and Siegel (1965) devised an instrument to measure 'educational set' on a continuum from the 'factually set' learner (predisposed to learn factual content) to the 'conceptually set' learner (preferring to learn concepts and principles). They comment that "'educational set' as measured ... probably reflects the learner's broad cognitive style rather than a more narrow perceptual process" (p.11), and find that these cognitive styles interact with the teacher's subject-matter emphasis:

> Conceptually set subjects were adversely affected by emphasizing factual rather than conceptual acquisition; conversely, factually set subjects were adversely affected by emphasizing conceptual rather than factual acquisition. (p.10)

Coop and Brown (1970) using the analytic versus non-analytic cognitive style construct defined by Kagan, Moss, and Sigel (1963) examined the effect of cognitive style and instructional treatment on factual content achievement, conceptualization-generalization content and total content. They found that 'teacher structured presentation' was superior to 'independent problem solving' on all three measures. However, they found no significant difference between students with an analytic cognitive style and students with a non-analytic cognitive style on any of the dependent measures, and found no significant interaction between cognitive style and instructional treatment. They suggest that the latter finding may have been due, in part, to the fact that the teaching methods were not specifically designed to contrast in a way that was congruent with the contrasting cognitive styles.

Gray and Kneif (1975) also used measures of cognitive style developed by Kagan, Moss, and Sigel (1963) in a main-effects study and found low correlations between various cognitive style dimensions and achievement. Of greater interest to this study, however, is their post-hoc analysis of intact classrooms which they claim indicates that the relationship between cognitive style and school achievement changed from one classroom setting to another.

Walters and Sieben (1974) using measures of field-dependence/independence (Witkin, Dyk, Faterson, Goodenough, & Karp, 1962) found that children at the field-dependent (global) end of the cognitive style continuum were significantly less well-equipped, both in terms of attitudes towards learning science in an open-ended environment and in terms of cognitive processes. They conclude:

> It is becoming evident that all children do not learn science the same way. The 'free and unguided exploration' approach may well be appropriate for some students, but may be very inappropriate for others. (p.73)

A recent study by Tamir (1975) reports several interactions involving cognitive preferences (Heath, 1964): for example between curricula (BSCS and non-BSCS) and school environments (city schools and kibbutz schools); between teachers' curricular bias (BSCS-supporters and non-supporters) and school environments; between curricula and teachers' curricular bias. He also concludes that cognitive preferences are partially dependent on the subject-matter and may have important uses in curriculum evaluation.

These few examples and other studies discussed in Chapter 5 strongly suggest that cognitive style variables should be considered in research on teaching and learning. With the exception of Coop and Brown's study the studies mentioned above suggest that cognitive style may interact with instructional treatment.

3.6 PRIOR ACHIEVEMENT AND INSTRUCTIONAL TREATMENT

In his recent review of what he calls the 'alternative abilities' model of ATI, Tobias (1976) isolates several problems. Firstly, he suggests that alternative instructional treatments designed to exploit different abilities may not rely on these abilities exclusively. It may be that an instructional treatment X designed to capitalize on ability x may be equally well coped with by ability y, for which treatment Y was specially designed, or both treatments X and Y may be effectively coped with by a third ability z for which neither of them was designed. In such an event interaction effects would not be evident. Secondly, he points out that it has been demonstrated that the abilities required by a particular task may alter as the learner engages in the task, and that treatment may need to take this into account. Thirdly, he quotes Nuthall (1968) and Tallmadge and Shearer (1971) in suggesting that ATIs may not be readily generalizable to different content areas. Finally, he doubts the practicability of designing alternative instructional treatments relying on independent abilities which last for a significant length of time in a classroom teaching situation. However, he refers to a number of studies which indicate consistent interactions between prior achievement and the degree of instructional support: the higher the level of prior achievement, the lower the instructional support required to accomplish instructional objectives. In Ausubel's terms (1968) one variable influencing learning and retention is the availability in the cognitive structure of specifically relevant anchoring ideas to bring about subsumption of the new idea.

Among the studies cited by Tobias is that of Abramson and Kagen (1975) who familiarized half of their subjects with the verbal and pictorial content of the instructional programs before randomly assigning them to a program giving low instructional support, or one giving high instructional support. They found that familiarization improved the performance of those assigned to the low support program, but reduced the performance of those assigned to the high support program. A further analysis showed that familiarization improved the performance of students of high average IQ assigned to both programs, but that the disordinal interaction mentioned above occurred with students of even higher (superior) IQ. A decreased performance on the part of an able group under conditions of high instructional support has led Tobias, and Cronbach and Snow (1969) to suggest that boredom and lack of attentiveness are responsible. A study by Tobias and Ingber (1976) specifically tested the hypothesis of interaction between prior achievement and instructional support:

> As predicted, an interaction between pretest score and instructional support was found, in which the superiority of constructed responding [high support] compared to reading [low support] decreased with increases in pretest score. (Tobias, 1976, p.70)

Mayer (1975) also reports significant interactions between instructional treatment and students' possession of specifically relevant knowledge, as would be predicted by Ausubel's (1968) view of the role of anchoring ideas in the subsumption of new knowledge.

Prior achievement has, of course, often been used as a basis for classroom grouping, with some students assigned to classes with an enriched program or an accelerated pace, and others to programs taken at a slower pace. The interactive effects quoted above, if sustained by further research, seem to suggest that although maximum support is beneficial to students with low achievement or lack of familiarity in a given area, it may indeed be counter-productive with students of high prior achievement or familiarity.

3.7 CONCLUSION

The earlier reviews of Bracht and Glass (1968), Bracht (1970), and Cronbach and Snow (1969) concluded that the empirical evidence has demonstrated few if any ATI effects. However, Cronbach and Snow (1969) and Berliner and Cahen (1973), even in their caution, find it impossible to believe that students do not differ with respect to the instructional treatment which suits each one best. Mitchell (1969) and Glaser (1972) are much more optimistic about the prospects of discovering significant disordinal ATIs, and Hunt (1975), Daniels and Stevens (1976), and Tobias (1976) present supportive evidence.

Cronbach and Snow are led to conclude that information-processing characteristics may prove more interactive than the traditionally researched abilities, and that differential effects may result more readily from treatments which differ in the degree to which responsibility is placed on the learner to organize his own material. These conclusions receive substantial support from the other research studies discussed.

It is noted that in those few studies which report interaction between general ability and treatment, the treatment which is beneficial to students of lower ability gives high instructional support, and may function through providing information-processing functions which the students do not themselves possess.

There is only limited evidence of interaction of instructional treatment with special abilities, and there appears to be no theoretical basis for linking special abilities with the assumed characteristics of the contrasting instructional treatments in this study.

There appears to be evidence of interaction of anxiety with task difficulty, and support for the hypothesis that where the learning task requires only one response students with high drive should perform well, but where the task allows for several competing responses a high level of drive may be disadvantageous. Although some carefully conducted studies find no evidence of disordinal interaction, others have found evidence of treatment interactions with personality characteristics such as the Ai scale of the CPI, Ascendancy and Responsibility from the Gordon Personal Profile and scores on the Scientific Interest scale of the Kuder Vocational Preference Record. Recent investigations attempting to match features of the instructional treatment with variations in internal/external locus of control, and conceptual level have produced strong disordinal interactions.

There are indications that elements of cognitive style, such as field-dependence/independence and Heath's (1964) cognitive preferences interact with instructional treatment, and in particular with the free and unguided exploration approach sometimes found in science teaching.

Finally it is suggested that high instructional support is beneficial to students with low prior achievement or familiarity with the subject matter, but is counter-productive with students of high prior achievement or familiarity.

CHAPTER 4

THE DESIGN OF THE INVESTIGATION

4.1 CONTRASTING INSTRUCTIONAL TREATMENTS

4.2 ATTRIBUTE-TREATMENT INTERACTION

4.3 RESEARCH IN THE CLASSROOM

4.4 OUTLINE OF THE DESIGN

  Setting

  Treatment

  Experimental Population

  Criterion Tests

  The Analysis

4.1 CONTRASTING INSTRUCTIONAL TREATMENTS

The development of inquiry skills has been an important aim of science curricula developed since the 1960's and a number of teaching strategies have been widely advocated as being more effective than traditional strategies in achieving this aim. However, the claims for the role of these strategies in the acquisition of the inquiry skills do not appear to be substantiated by the research reviewed in Chapter 2.

Many of the studies reviewed have suffered from an inadequate description of the alternative treatments (Coombs, 1966). Treatments may not live up to the promise of their descriptions. There is, for example, considerable suspicion that, in many studies contrasting laboratory strategies with lecture-discussion strategies, the laboratory has been used in a fashion inimical to the development of inquiry skills (Herron, 1971). Also, the same teaching strategy utilized by two different teachers may differ more significantly than examples of the so-called 'contrasting' strategies (Egelston, 1973). A comparison of the studies of Boleratz (1967) and Ripple, Millman, and Glock (1969) indicates that care must be exercised so that the examples of the contrasting strategies are equally good examples of those strategies. Appelations such as 'traditional lecture-discussion', 'didactic', and 'teacher-centred' in the research literature tend to carry pejorative overtones, and it is suspected that poor teaching is thought to be typical of these strategies.

The nature of the criterion tests is also of substantial importance (Thomas & Snider, 1969). An unfair comparison is made of alternative strategies if the strategies are not claimed to be directed to the same outcomes. One of the consistent findings of the research on the use of the laboratory in science teaching is that students who use the laboratory develop certain laboratory skills: it is only the claim that use of the laboratory develops inquiry skills that is unsubstantiated. An innovative strategy is likely to have new or enlarged purposes and is not likely to be shown to be more effective than traditional strategies as shown by tests measuring the traditional outcomes. However, a test designed to measure the newly emphasized outcomes is likely to show that the traditional strategies are less effective. Another illustration of the need for content validity in criterion tests is given by Tamir and Glassman (1970, 1971a, 1971b) who observed that tests of manipulative skills and such tests as the Watson-Glaser Critical Thinking Appraisal did not adequately sample the objectives of laboratory work in BSCS courses.

Even when there are no serious reservations on the methodological issues just outlined, research studies do not appear to give unequivocal support to the claims of inductive teaching strategies over deductive strategies, nor of guided discovery strategies over didactic strategies (Shulman & Keislar, 1966; Atkin & Burnett, 1969; Belanger, 1969; Craig, 1969; Hermann, 1969; Ramsey & Howe, 1969a, 1969b; Novak, Ring, & Tamir, 1971; Shulman & Tamir, 1973). But there are suggestions that alternative strategies suit students with different personality characteristics, skills, and abilities (Thomas & Snider, 1969; Berliner & Cahen, 1973; Sakmyser, 1974).

4.2 ATTRIBUTE-TREATMENT INTERACTION

Despite the wide intuitive support for the investigation of attribute-treatment interactions, the early research was inconclusive and the reviews of Bracht and Glass (1968) and Bracht (1970) held out little hope for the success of this line of investigation. However, practitioners, researchers, and reviewers alike continue to insist that students who differ in all sorts of other ways also differ in their susceptibility to particular teaching methods. The more recent reviews of Hunt (1975) and Tobias (1976) are distinctly more hopeful in tone, and the studies reviewed in Chapter 3 have encouraged the present investigation.

One promising line of investigation is embodied in the suggestion of Glaser (1972) that researchers should attempt to discover and refine 'new aptitudes' which describe students' differential susceptibility to variations in the instructional environment. This view accords with a general conclusion of Cronbach and Snow's (1969) comprehensive review that information-processing attributes may prove to be more interactive with instructional treatments than the traditionally researched abilities. There are several cognitive processes which have been identified experimentally and there is some evidence of interaction between these process variables and instructional treatment (Siegel & Siegel, 1965; Walters & Sieben, 1974; Gray & Kneif, 1975; Tamir, 1975).

A recurring theme in reviews, both of research into contrasting instructional treatments and ATI research, is the observation that many of the studies suffered from the brevity and artificiality of the treatment (Cronbach, 1966; Wittrock, 1968; Cronbach & Snow, 1969). In the following section these issues will be considered in more detail.

4.3 RESEARCH IN THE CLASSROOM

There appear to be parallels between the forms of research undertaken in education and in the biological sciences. The earliest accumulation of knowledge in biology resulted from the distillation of the insights of practitioners attempting to elucidate the meaning of their own observation and experience. Problems tended to be looked at in a holistic fashion and many valuable insights were gained, but such investigations tended to be non-cumulative and unsystematic. This is the practice of the naturalist who, in education, has his parallel in the articulate classroom practitioner.

The development of the scientific tradition with its emphasis on the control of extraneous independent variables and explicit and logically-defensible, if not always impeccable, statistical analysis, appeared to promise much more. There is no doubt about the value, both in terms of volume and significance, of the knowledge resulting from this tradition. The greatest impact of this tradition has been felt in the pure and applied physical sciences where variables can be described with a high degree of accuracy and controlled with a high degree of certainty. In sciences where living material is the subject of study, description and control are more difficult. Living material itself is an extraordinarily complex physico-chemical system in dynamic equilibrium and is possessed of incalculable variability as the result of the unique interaction of genetic recombination and environmental influences. Nonetheless, the techniques of experimental methodology and statistical analysis developed in the fields of behavioural science and biological science have also added enormously to knowledge and insight. The physiologist has his parallel in the experimental psychologist.

Although the contribution of physiologists, biochemists, experimental psychologists and the like tell us much about the behaviour of living organisms, the study of the natural behaviour of organisms in their ecological setting has called for different methodologies that characterize ethology. Ethological studies can only be pursued in the natural multivariate environment over large periods of time. Similarly, the tradition of laboratory experimentation, with its emphasis on the control of extraneous independent variables, has not been noticeably fruitful in elucidating the complexities of natural classroom behaviour. It is argued that education needs researchers like the ethologist to carry out investigations which have ecological validity. Such investigations would have characteristics in common with the traditions of both the reflective analyst and the experimentalist: the approach would be holistic and naturalistic, and the methodology would be systematic and lead to valid generalizations.

A development of the concept of representativeness or ecological validity in experimental design can be traced through such influential papers as Campbell and Stanley (1963), Bracht and Glass (1968), Snow (1968), Pereboom (1971), Snow (1974), and Bronfenbrenner (1976).

Campbell and Stanley (1963) point out in their opening remarks that in research on teaching the experimenter is working with an "intransigent" environment and cannot exercise the complete control characteristic of experimental designs in the tradition of Fisher (1925). They add that the "most conspicuous effect [of the Fisher tradition] seems to have been to elaborate statistical analysis rather than to aid in securing 'adequate and proper data'" (p.172). Fisher's strategy of the pre-experimental equation of treatment groups through random assignment can seldom be achieved when working in schools, and the majority of the experimental designs for research on teaching discussed by Campbell and Stanley are designated by them as quasi-experimental designs.

In referring to the interpretation of experimental results

Campbell and Stanley distinguish between internal validity (did the experimental treatments make a difference in this instance?) and external validity or representativeness (to what populations, settings, treatment and measurement variables can this effect be generalized?). In discussing various research designs they comment on the various threats to the internal and external validity of the designs. Threats to representativeness are seen to be the reactive effect of pretesting which may affect a respondent's sensitivity to the treatment, interaction between the treatment and any bias in the selection of the comparison groups, the reactive effects of the experimental arrangements, and interference that may result from prior treatments.

Bracht and Glass (1968) have elaborated on the threats to external validity and have illustrated additional threats. They distinguish between two main types of external validity: population validity (what population can be expected to behave in the same way as the experimental population?) and ecological validity (under what settings, treatments, dependent variables, etc., can the same results be expected?).

In generalizing from the experimental environment to other ecological settings there are a number of problems which may result from the experimental setting not being entirely typical or representative, and these stem in a large measure from the fact that an experiment or a survey is being carried out at all; the experimenter and his procedures intrude into the life activities of the classroom. There are problems in knowing what outcomes are expected from the treatments and in knowing what the criterion tests measure. Then there is the problem of describing explicitly what the experimental treatment is, and of the possibilities of confounding by the interaction of the treatment and events extraneous to the experimental environment. And herein lies a dilemma. To what extent should attempts be made to control extraneous independent variables?

The methodological thrust of experimental psychology is towards tight control of the experimental environment and the exclusion of, or control for, extraneous independent variables. However, does not the exclusion of, or control for, such variables make the experiment unrepresentative and limit its generalizability to the 'unreal' laboratory environment? In other words, what is 'extraneous'? Brunswik (1956) argues for the study of subjects in their natural ecological settings:

> Proper sampling of situations and problems may in the end be more important than proper sampling of subjects, considering the fact that individuals are probably on the whole much more alike than are situations among one another. (p.39)

Millman (1966) has also stressed that an experiment should be fully representative of the variety of conditions to which it may be intended to generalize the results. Cattell (1966) argues against manipulative experimentation on the grounds that it risks disturbing the very processes which are to be observed. And McKeachie (1974) argues that part of the reason the laws of learning are often found not to hold in educational situations is that there has been a "failure to take account of important variables controlled in laboratory situations but interacting with independent variables in natural educational settings" (p.9).

Pereboom (1971) argues that an experimental laboratory approach to a multivariate discipline has significant problems in control and in generality, and comments on the value of naturalistic observation:

> The behavior of an animal out of his ecological niche may be ... as enigmatic as a piece of a jigsaw puzzle out of *its* appropriate niche. Only when knowledge of the behavior of the naturalistically-evolving organism ... in his appropriate bio-environment is available, would we then be able to have any insight into such laboratory phenomena. (It might also be noted that a greater degree of behavioral stability may take place in the natural setting than in the lab, possibly due to the greater redundancy of relevant cues, both spatially and temporally.) (p.445)

and again, "if one wishes generality as well as naturalistic relevance, one's efforts might be better directed toward optimizing observation and recording, and minimizing control and manipulation" (p.452).

Snow's (1974) important paper can be seen as a direct development and extension of the explorations on research design by Campbell and Stanley (1963) and Bracht and Glass (1968), and was strongly influenced by an approach to experimentation proposed by Brunswik (1956) (see Snow, 1968). In observing that the greatest threat to the external validity of educational experiments occurs when the experiment does not fit the nature of the behaviour being studied, he recalls Brunswik's (1956) suggestion that we should continue to use systematic design for what it is good for, but that it is not good for the adaptive probabilistic functioning of human behaviour that is found in the natural school setting.

In discussing population validity Snow points out that the generalization of findings from a sample population to the accessible population and then to the target population depends on an assumption of random sampling that can seldom be achieved. However, one can at least generalize from a sample to a hypothetical accessible population comprising individuals like those observed (Lindquist, 1953; Cornfield & Tukey, 1956; Bracht & Glass, 1968). The key to generalization to the target population is a detailed knowledge and description of the characteristics of the populations, and by the testing of interactions in the sample population which will limit the generalization of treatment effects. It is just not feasible to measure all possible student characteristics and so the investigator includes those attributes that theory, prior research, and the investigator's experience and intuition suggest are relevant to the treatment and dependent variables. In these circumstances Snow suggests that the study will be quasi-representative with respect to the population characteristics.

Snow advances basically the same argument with respect to ecological validity. Ecological validity depends on the extent to which the environment and situations compared in an experiment are representative of those situations to which the experimenter wishes to generalize. Snow observes:

> A truly representative design with respect to ecological validity would sample randomly from the universe of treatments that is the target for future generalization. Since this is usually impossible, the recourse is again to detailed description of the universes and sample treatments at hand and the testing of interactions. (p.272)

In such a representative design, treatment characteristics can be assumed to be multivariate and interrelated, and Snow suggests that quasi-representative designs require "methods of analysis that allow those characteristics to covary as they do in nature" (p.272).

Snow (1974) suggests the following steps towards representativeness

1. Experiments should be embedded in their relevant contexts. If it is proposed to generalize to other classrooms the investigation should be set as unobtrusively as possible into the normal classroom environment and programme.
2. By conducting an experiment anew with different subjects increased population representativeness should be gained, and by varying the teachers, methods, content, or context

increased ecological representativeness should be gained. However, there are many reviews of educational research where roughly the same experimental comparison has been tested in a number of independent studies with inconclusive results (for example, see reviews on discovery learning mentioned in Chapter 2).

Snow gives additional examples including a review of 212 replications contained in a study of performance contracting vs. conventional teaching: there were 124 non-significant differences, 28 significant differences favouring performance contracts and 60 significant differences favouring conventional teaching. He observes mildly,"apparently, replication by itself does not produce generalization" (p.277).

3. Observation of the subjects during the experiment can give valuable information on their perception of the conditions of the experiment and on their reactions to these conditions.
4. Observation of the contextual setting of the experiment should give an indication of untoward effects that may reduce the ecological representativeness of the study: such effects as novelty and Hawthorne effects or unusual events in the history or maturation of the subjects.
5. The inclusion of aptitude variables and the testing of interactions reflect on population validity and hence the generalizability of the experiment. Snow suggests that even when measures of special aptitudes that relate specifically to the experimental treatments are available, measures of general ability should always be

included to secure greater referent generality. And in referring to the testing of interactions Snow comments:

> Blocking continuous variables loses information and weakens the analysis. Furthermore such variables are likely to be correlated in the population; even when equal cell size is achieved, their effects will be misrepresented in conventional analyses of variance. (p.280)

6. If newly-invented or unusual instructional treatments are used greater representativeness would follow from the preparation of students so that their behaviour in response to the treatments is as "smooth and habitual as in ordinary school learning" (p.280).
7. Most school programs are extensive in time and valid generalizations about school learning outcomes should be based on investigations using substantial samples of learning time. A lengthy treatment does open the investigation more to the effects of extraneous variables such as those affecting the history and maturation of students but Snow (1974) and Pereboom (1971) both view control as being less important than representativeness. Where, as in the present investigation, natural teaching styles are the contrasting instructional treatments, there is need to allow the teachers to exercise their full repertoire of teaching behaviour over a substantial period to achieve a stable reflection of the frequencies of such behaviour. It is the experience of practising teachers, on meeting new classes at the beginning of the school year, that it takes a considerable time to establish what might be regarded as a stable pattern of teacher-pupil interaction.
8. In classroom research involving the comparison of

instructional treatments there may be severe difficulties in fitting an imposed teaching strategy to the natural instructional styles of the teachers involved and generalization may be limited by the confounding of teacher and an imposed method. One of Snow's suggestions is that systematic comparisons be made of complex treatments and that the key mechanisms be identified by the method of successive omission of factors.

9. In his discussion of the population and ecological validity of experiments Snow also refers to an experiment's 'referent generality': its external value in understanding the phenomenon being studied. One indication of referent generality is the influence the experimental treatment has on "fairly general, enduring, or pervasive characteristics" (p.274) of subjects. This suggests that among the outcomes of an experiment these more general or pervasive characteristics should be measured.

Bronfenbrenner (1976) distinguishes between the 'contrived experiment' of the laboratory and the 'natural experiment' set in its real-life educational setting and gives his support to the latter:

> Our researches cannot be restricted to the laboratory; for the most part, they must be carried out in real-life educational settings .... This does not mean that laboratory experiments cannot serve a useful and, indeed, essential purpose, but they must be carried out with explicit recognition of the delimiting and distorting nature of the laboratory as a setting and deliberately designed to articulate closely with and complement companion researches carried out in real-life situations. (p.5)

He points out that the methodological objective of both the contrived and natural experiment are the same: to set two similar but different systems side by side in order to see more clearly the differences between them. However he stresses that,

> the primary purpose of the ecological experiment becomes not hypothesis-testing but discovery - the identification of these systems-properties and processes that affect, and are affected by, the behavior and development of the learner. (p.6)

Bronfenbrenner describes the ecological structure of the educational environment in considerable detail, placing the immediate educational setting of the classroom in the progressively wider contexts of the neighbourhood, society, and culture. He accepts that in the ecological model he adopts, person-environment and environment-environment relationships are dynamic and so the experimental strategies that follow would be necessarily complex. This present investigation is however much less ambitious and seeks to discover relationships that exist within the classroom setting. It is recognized that the wider community and social contexts in which these classrooms are set will have an influence, but, in short, it is contended that they are representative social contexts that add to the 'contextual validity' desired by Bronfenbrenner (1976, p.7).

In considering the properties of the immediate educational setting and their implications for research design Bronfenbrenner states that to secure 'ecological validity' research participants should not be expected to work in an alien social setting, nor in roles or activities which do not have an established social meaning for them. He also asserts that "the experiment must also last long enough to approximate what happens in real life" (p.7). To preserve the 'ecological integrity' of the setting a determined effort has to be made to keep experimental intrusion to a minimum.

In moving to 'natural experiments', the pervasive tradition of the laboratory, and its efforts to eschew 'subjective' data, can still blinker the researcher. Bronfenbrenner comments on two common omissions in naturalistic experiments, the failure to provide an adequate description of the natural environment and the participants, and the failure to take account of the participants' perception of the environment. The present investigation is deficient in the latter respect and may well have included an instrument similar to Tuckman's (1970) Student Perception of Teaching Style, but based on the distinctions made in the category system used in this research for analyzing teaching style. The work of Tisher and Power (1975) emphasizes the importance of students' perceptions of their learning environment (see also Tisher, 1977). It can be seen to be clearly important to check whether the experimental treatments are seen by the participants in the same way as conceptualized and described by the experimenter, and Bronfenbrenner refers to this aspect of research strategy as checking on its 'phenomenological validity'.

4.4 OUTLINE OF THE DESIGN

The purpose of the present investigation is to elucidate the relationships that exist between selected student characteristics and contrasting styles of teaching biology in secondary schools as reflected in measures of achievement and attitude. The argument marshalled above lends support to a research strategy with the following characteristics. These brief descriptions are amplified in the chapters which follow.

Setting

In Bronfenbrenner's usage, "the factors of place, time, activity, and role constitute the elements of a setting" (1976, p.5). In one sense the setting is but one aspect of the treatment variables, but it is important in the interests of generalizability to look specifically at the context in which the instructional treatments are applied.

The investigation was carried out in eight Grade 11 classes taking the Web of Life biology course. The investigation spanned the full school year and required no alteration to the normal school organization or the teachers' conduct of their classes. The experimental intrusion was limited to the administration of tests of general ability, personality characteristics, and cognitive preferences in the first two weeks of the year before formal teaching had begun, and periodic, and unannounced, visits during the year to observe and record classroom behaviour. The achievement tests and attitude inventory were administered by the teachers within their regular schedule of testing. The basic intention of these strategies was to preserve, as far as possible, the natural classroom situation, representative of biology teaching in Victorian schools.

As argued above in section 4.3, the control of extraneous independent variables was largely ignored in the interests of ecological validity. However, since some classes were to be regarded as receiving one kind of instructional treatment and other classes the alternative treatment, it was important that the schools themselves did not vary significantly in terms of their general sociological setting. That is, it was important to exercise some control over these extraneous variables by homogeneity. All schools were co-educational and situated in the eastern and south-eastern districts of Melbourne which drew from areas of similar socioeconomic status. No high-fee non-government schools were included. The one non-government school in the sample was a Catholic regional college: a systemic co-educational school with a predominantly middle-class intake and provisions similar to the government schools.

Although a church school, it had a predominantly lay teaching staff and, in the biology class at least, observation provided no evidence of an ethos which might have distinguished it from other schools in the sample.

Treatment

As detailed in Chapter 6, eight teachers were chosen who on the best evidence available were recognized as good teachers and who, while teaching the same course content and intending to follow the teaching strategies recommended in the teaching materials towards the same agreed ends, would nevertheless fall into two groups exhibiting contrasting natural teaching styles. It was expected that four teachers would fall in each group. An extreme-groups design was chosen (Flanders, Note 2) as a relatively powerful way of establishing interactions (Cronbach & Snow, 1969, p.21). However, utilization of natural teaching styles in the interests of ecological validity preclude random sampling from the experimentally accessible population, or random assignment to treatment groups. So the conditions for the use of inferential statistics in generalizing to the accessible population are not fulfilled. As discussed above in section 4.3, the most appropriate procedure is then to describe the instructional treatments as precisely as possible and to limit generalization to a hypothetical accessible population like the sample.

The classroom behaviour of the eight teachers was observed and recorded periodically throughout the year but, apart from the initial pretesting and this periodic observation, there was no interference either in the organization of their teaching program or to their teaching methods. Utilizing the teachers' natural styles over an extensive period should allow a stable pattern of teacher and pupil behaviour to develop and should provide the opportunity for a wide sampling of classroom situations. The observation and recording techniques provide the basis for a detailed description of the treatment variables, and some check on untoward events which might threaten representativeness. The extensive period leads to less control but to greater representativeness, and the reactive effects of the pretesting at the beginning of the year can, in the natural ebb and flow of a year's program, be assumed to be minimal.

The audiotape recordings and concurrent anecdotal record made of each lesson observed were converted into an annotated transcript This was then analyzed by a multiple-coding category system to provide a quantitative and, to some extent, a qualitative record of the teachers' classroom behaviour. That the teachers fell into the expected contrasting groups was tested by carrying out two forms of cluster analysis on this data: an 'inverted' factor analysis (Stephenson, 1936, 1953) and the H group method (Veldman, 1967). Both analyses indicated that there were in fact two contrasting groups but that, contrary to expectations, three teachers belonged to one group and five to the other. The fact that the treatment groups were not validated until after the treatment had been completed meant that there was no certainty at the beginning of the investigation as to which treatment group a particular class belonged. This was some protection against experimental bias or Hawthorne effects. The description and validation of the contrasting treatments is an important element of the investigation and is described in Chapter 6.

Experimental Population

The experimental population consisted of students in the 11th grade biology class groups of the eight teachers chosen for investigation. The present investigation was set in 11th grade where biology is studied as a subject for the first time, so that prior achievement in, or attitudes towards, biology would not be likely to affect the outcomes. However, among the 172 students from whom data on student attributes were originally collected, there were 11 who were repeating 11th grade biology. It could not be known whether the achievement or attitude of these students was due to their previous teaching in the subject, or to their present teaching, so that although all data were collected for these students they were not used in the analysis. There were no more than two repeat students in any one class, and it was believed that their presence in the class and their absence from the analysis would not significantly affect the conclusions drawn from the study.

In addition there were 15 students who were present when the data on student attributes were collected, but were not there when the attitude inventory or achievement tests were administered at the end of the year: five had transferred to other schools, eight had left school to enter the work force, and two were absent from the final testing program. There were no additions to the classes during the year. Correlational analysis depends on complete data, and the data on these 15 students were not used. The experimental population consisted of the 146 11th grade biology students taking biology for the first time on whom complete data were available.

Intact class groups are used not only to minimize disruption to schools, and to the students, but to preserve the ecological validity of the study so that findings may be generalized to similar natural school settings. However, in the analysis of the data individuals were used as the unit of analysis rather than class means. Although "nearly all educational investigators have employed persons, rather than the class as the unit" (Cronbach & Snow, 1969, p.96), this strategy has often been criticised (Campbell & Stanley, 1963; Lumsdaine, 1963; Raths, 1967). As Peckham, Glass, and Hopkins (1969) observe, the fundamental difficulty is that observations made on students within a class may not be independent. Because of the moderating influences of classroom interaction, students in a class may become more similar in performance than students undergoing individual instructional treatment. Also a disruption to the normal class routine affects all students in that class but no students in other intact class groupings. The ANOVA model assumes the independence of observations, usually achieved by random assignment of individuals to treatment, and the use of individuals' scores within intact classes increases the likelihood of Type I errors.

If instructional treatment was the only independent variable, and if it were intended to make generalizations about teachers, then the class means would appear to be the appropriate data for analysis. However, the present investigation seeks to examine the interactions of student attributes (for which class means would be meaningless) with contrasting teaching styles as measured in student attitudes and achievement. And as Peckham et al. (1969) observe,

> when one uses group means rather than the individual subjects as the experimental unit, he is deprived of the opportunity of stratifying on certain organismic and demographic variables and, hence, loses all chance of detecting significant effects due to those sources or due to their interaction with other main effects. (p.344)

So that it would appear to be necessary to use the individual as the unit of analysis, but to be aware of the possibility of Type I errors.

Lord (1963) in discussing the random assignment of individuals to alternate treatments says:

> If the individuals are not assigned to the two treatments at random, then it is not too helpful to demonstrate statistically that the groups after treatment show more difference than would be expected from random assignment - unless, of course, the experimenter has special information showing that the non-random assignment was nevertheless random in effect. (p.38)

This is interpreted as meaning that if the investigator can establish that the groups have similar characteristics as they would if they had been generated through random assignment, then he may be justified in using statistical techniques that are based on the assumption of random assignment.

After validation of the two treatment groups by cluster analysis of their teachers (the groups were found to contain 67 and 79 students respectively) the groups were compared on the 13 attribute variables used in the study. As can be seen in Table 4.4.1, they differed on only one at the $\alpha = 0.05$ level. (On the basis of probability alone it is to be expected that of 20 such comparisons one will be significant at the $\alpha = 0.05$ level). In addition there is an inconsistent pattern of differences: three favouring one group and 10 the other.

However, before asserting that there is no significant difference between the groups, it is important to calculate the power of the t-test with sample sizes of $N_1 = 79$ and $N_2 = 67$. Cohen (1969) gives operational definitions of 'large' effects (0.8σ), 'medium' effects (0.5σ), and 'small' effects (0.2σ). For example, Penick and Brewer (1972) found that the average power of over 500 statistical tests reported in the Journal of Research in Science Teaching (1969-71) was 0.87 for a large effect, 0.71 for a medium effect, and 0.22 for a small effect. For research reported in the American Educational Research Journal (1969-71), Brewer (1972) found the corresponding values to be 0.78, 0.58, and 0.14. At an $\alpha = 0.05$ level, the

TABLE 4.4.1

t-TESTS FOR TREATMENT GROUPS ON THIRTEEN ATTRIBUTE VARIABLES

Variable	M_1	SD_1	M_2	SD_2	t
TI	19.10	7.16	19.58	5.37	-0.46
TO	16.84	5.30	17.58	4.64	-0.89
Es	10.18	4.09	10.55	4.09	-0.54
Co	15.61	4.14	15.75	4.51	-0.19
Au	20.80	7.21	23.75	6.11	-2.66*
RO	14.89	4.78	14.51	4.99	0.46
SE	19.81	6.67	20.21	6.29	-0.37
IE	29.62	8.59	31.27	9.78	-1.06
PI	26.44	9.14	24.81	8.59	1.10
AL	12.01	3.87	11.05	3.81	1.50
A/Q	19.76	3.44	19.91	3.55	-0.26
M/P	18.54	3.11	18.58	3.23	-0.08
B4OX	22.27	8.24	24.18	7.92	-1.41

$N_1 = 79$; $N_2 = 67$

* $p < 0.05$

Critical t(144) = 1.98 at the 0.05 level of confidence t-tests in Table 4.4.1 have power of 0.99, 0.85, 0.22 for Cohen's three effect levels. So that, in addition to finding that only one t-test out of 13 reached statistical significance at the $\alpha = 0.05$ level, it is known that, at this α level, the t-test has a probability of 0.99 of detecting a large (0.8σ) difference, a probability of 0.85 of detecting a medium (0.5σ) difference, and a probability of 0.22 of detecting a small (0.2σ) difference.

This indication of approximate equality between the groups is reassuring, in that prior manipulation of individuals or groups was eschewed in this investigation in favour of natural assemblages of participants and post-experimental validation of the treatment groups. And the use of the individual as the unit of analysis, necessitated by the desire to test the effects of personological variables and attribute-treatment interactions, may be less of a problem than is sometimes suggested (Herron, Luce, & Neie, 1976). In an author's reply to West (1977), Herron et al. comment further:

> When one may assume that assignment of experimental unit (be it individual or class) is random in effect, the results of an analysis which assumes the individual as the experimental unit is likely to be comparable to results from an analysis which assumes the class as the experimental unit, regardless of whether the treatment is in fact administered to the class as an intact unit or to individuals. (p.165)

Criterion tests

The two achievement tests used in this investigation were developed by the Australian Council for Educational Research (ACER), as described in Chapter 7, specifically for the Web of Life biology course. The tests can be assumed to have a high degree of content validity, and as all teachers were following the same course towards the same stated objectives there should have been no bias towards one or other of the treatment groups.

The attitudinal outcomes were measured by the Biology Attitude Inventory (BAI) (Lucas & Broadhurst, 1972) which was a modified version of an earlier attitude inventory developed for secondary school physics, The Physics Attitude Index (PAI) (Mackay, 1970, 1971a; Gardner, 1972). The BAI contains four scales measuring students' views about methods of teaching biology, views about biology as a process, views about scientists, and enjoyment of biology. These scales represent varying degrees of referent generality towards the 'distal' end of Snow's continuum (1974, p.275). The inclusion of measures of multiple outcomes of the investigation, including dependent variables with high referent generality, meets another of Snow's (1974) dicta concerning the representativeness of instructional experiments.

The analysis

The aim of the analysis is to discover relationships, especially attribute-treatment interactions if they occur, among the 6 dependent variables and 14 independent variables.

The multivariate nature of the investigation calls for multivariate statistics (Tatsuoka, 1969) and as described in Chapter 8, multiple regression procedures were used to analyze the relationships between each dependent variable and the set of prediction variables. A multiple regression approach is appropriate to the purpose of the investigation; it permits the analysis of multiple independent variables some of which may be related and covary in nature; it can deal with both categorical and continuous variables; there is no need to block continuous variables which would result in a loss of precision in measurement and a loss in degrees of freedom; and the relationships can be analyzed by the successive omission of factors from a more complete description as recommended by Snow

(1974, p.283).

In the multiple regression analysis used in this investigation the contribution of each variable to the explained variance was considered in turn, controlling for all others. Intelligence was the last variable to be dropped from the full regression equation, so that in addition to being a predictor variable in its own right it acts as a covariate for all other attribute variables. It is assumed in covariate analysis that individuals have been randomly assigned to treatment groups and that the covariate is independent of the treatment (Elashoff, 1969). In the present investigation it has been indicated that the allocation of individuals to treatment groups is random in effect, and all attribute variables were measured prior to the commencement of the treatment. The covariate analysis was regarded as an appropriate strategy and serves to increase the power of the experiment.

CHAPTER 5

STUDENT CHARACTERISTICS: ATTRIBUTES

5.1 INTELLIGENCE AND APTITUDES

5.2 PERSONALITY CHARACTERISTICS

5.3 MEASURES OF PERSONALITY

5.4 THE OMNIBUS PERSONALITY INVENTORY

5.5 COGNITIVE STYLE: COGNITIVE PREFERENCE

Students differ from one another on every conceivable physical and psychological dimension for which measurements have been made. The variability that exists in organisms, as a result of the unique interplay of the genetic encoding of living material and their environment, is a fundamental principle of biology. It is, however, their psychological variability that is the principal concern in education. Even within this more limited field it is clearly impossible in a study of this nature to consider, as separate variables, all the psychological constructs which have been investigated previously, so a small number of characteristics have been singled out for special study. The selection of these characteristics was based on a judgement as to what personological variables were likely to interact with teaching styles which contrasted in terms of their inquiry-centredness, or authority-centredness.

Tyler (1969), in her review article on individual differences, discusses several classes of student characteristics that have been the focus of educational research. These characteristics can be conveniently grouped as follows: achievement, intelligence and aptitudes, motivation and personality, and cognitive styles. In this study measures of achievement are dependent variables and the other three classes of personological variables are independent variables.

Let us now consider in turn the measures of each of these three classes of independent variables employed in this study.

5.1 INTELLIGENCE AND APTITUDES

Glaser (1972) points out that in our traditional educational settings, where there is minimal variation in the conditions under which children are expected to learn, the individual differences that are customarily measured in order to differentiate learning assignments centre around the concepts of intelligence and aptitudes. In fact, Binet's first attempts to measure intellectual ability were prompted by a desire to identify mentally-retarded children in the schools of Paris so that they might be given instruction more suited to their limited ability. Cronbach (1970), in considering the construct of intelligence, points out that if we judge intelligence tests on the basis of their predictive validity it is clear that they are, in effect, tests of scholastic aptitude or scholastic ability. On this ground alone there would seem to be no need to justify further the inclusion of a measure of general ability as a potentially fruitful independent variable in this investigation. Although the evidence is by no means unequivocal, the research reviewed in Chapter 3 suggests that there are possible interactions between general ability and instructional treatments which provide varying degrees of assistance in information-processing (Cronbach & Snow, 1969; Salomon, 1974).

Tests have been developed which provide measures on a variety of differential aptitudes such as spatial, mechanical, and abstract reasoning aptitudes. However, McNemar (1964) concludes that on the basis of validity coefficients quoted, multi-test differential aptitude batteries have not proved their superiority over tests of general intelligence as predictors of school achievement except in areas of numerical ability and school achievement in mathematics. And although there is some evidence of interaction between special aptitudes, such as mathematical aptitude, and contrasting instructional treatments, there do not appear to be any strong theoretical grounds for including measures of special aptitudes in the present study.

Some studies have also used measures of prior achievement as covariates of their criterion measures of achievement. Appropriate measures of prior achievement, however, were not available for this study. This investigation was set in the 11th grade where biology is studied as a subject in its own right for the first time in order to avoid the risk of prior attitudes towards, and achievement in, biology from affecting the relationships between instructional treatment and outcomes. Although all students could be presumed to have had a similar history of four years secondary school study in science, the various measures of achievement were unique to each school, and it was not possible to reduce them to a common scale. The same problem exists with other measures of prior school achievement.

It was decided therefore to use a measure of intelligence, or general ability, to represent the class of ability measures. One school had planned to administer the ACER Advanced Test B40 to its 11th grade students, and it was decided to take advantage of this testing program and to administer the same test to the other classes in the study. The ACER Advanced Test B40 is a group intelligence test of 77 items designed for the senior high school/college age range and standardized against the Otis Higher C test. The corrected split-half reliability is given as 0.89, and the product-moment correlation of the B40 with the Otis Higher C is given as 0.80 (ACER, n.d.). The raw scores on this test were used as measures of the students' general ability rather than intelligence quotients which are normed against a larger, and, for this group of 11th grade biology students, a not entirely appropriate population.

5.2 PERSONALITY CHARACTERISTICS

If one accepts Allport's (1963) definition of personality as "the dynamic organisation within the individual of those psychophysical systems that determine his characteristic behaviour and thought" (p.28), it appears obvious that personality will be an important influence in determining his reactions to learning situations in the classroom. In considering the individual differences among students that may interact with instructional treatment, Cronbach (1966) believed that "the interacting variables may have more to do with personality than with ability" (p.90).

Although Sells (1969) refers to the "disappointing experience with personality measures in prediction of academic performance" (p.943) reference is made below to some literature which suggests that aspects of personality may well interact with instructional treatment in determining both cognitive and affective outcomes.

Anxiety

Ruebush (1963) states that despite the various shades of meaning attributed to the construct of anxiety by investigators adopting particular theoretical orientations and operational criteria, there appears to be a common core of meaning. Anxiety is agreed to be a distinguishable unpleasant emotional state which may also have measurable physiological concomitants. However, Gaudry and Spielberger (1971) observe that in much of the research using anxiety scales it is not clear whether anxiety is being viewed as a relatively stable personality trait, or as a transitory emotional state. In the studies which are noted below anxiety is viewed as a motivational or drive-related construct, and considered as a predispositional state variable which can be heightened in situations of stress. Testing or evaluative situations are seen as specific instances of stressful situations. The student with a high level of anxiety may be characterized as being highly dependent on external support and being easily discouraged by adverse criticism.

The review of Ruebush (1963) further indicates the wide appeal that 'anxiety' has had for investigators of varying orientations. Levy (1961) suggests that more recent popularity may be connected with the development and availability of reliable instruments for measuring anxiety, such as those of Taylor (1953), Castaneda, McCandless, and Palermo (1956), Sarason, Davidson, Lighthall, Waite, and Ruebush (1960), and Cattell, Beloff, and Coan (1958). The studies reviewed by Ruebush indicate that anxiety either helps or hinders achievement depending on the students' perception of the testing situation and the difficulty of the task. In the description of their modification, for children, of Taylor's (1953) scale of manifest anxiety, Castaneda et al. (1956) report over 20 studies investigating the relationship between anxiety and various simple and complex learning tasks. In general, the studies suggest that for simple learning tasks, such as classical conditioning, students with high anxiety perform better than students with low anxiety, but for more complex learning tasks, such as trial and error learning, they do not perform as well.

Cox (1960) and Neville, Pfost, and Dobbs (1967) provide evidence to support the Yerkes-Dodson (1908) Law which would suggest that the relationship between anxiety and learning takes the form of an inverted U-shaped curve. Klein, Fredericksen, and Evans (1969) also find that test anxiety had a consistent curvilinear relationship with test performance, but in their case poor performance was associated with an intermediate level of anxiety. Gaudry and Spielberger (1971) cite other studies which find that the middle-anxiety group perform worst, so the Yerkes-Dodson principle should not be accepted incautiously.

Tallmadge and Shearer (1969) found that anxiety and introversion/extroversion were correlates of the two learning styles they postulated. Students who exhibit the first learning style (characterized by technological interests) tend to low anxiety and to introversion; students who exhibit the second learning style (characterized by social/aesthetic interests) tend to high anxiety and to extroversion. Grimes and Allinsmith (1961) suggest that the achievement of anxious students is, in part, a function of the degree of structuring of the learning task. Carlson and Ryan (1969), working with students at high school and college level, found significant negative correlations between test anxiety (Sarason, 1958) and performance on Knowledge and Comprehension (Bloom et al., 1956) tasks, but no relationship with higher levels of cognitive performance. They suggest that test anxiety interferes with task performance when the student is given no information to work with, but must rely on memory alone.

These studies, and those reviewed in section 3.4, suggest that the level of anxiety of a student may interact with the nature of the material and its mode of presentation. The student is thrown more onto his own resources and receives less support from the inquiry-centred teacher than the authority-centred teacher.

Dogmatism

The construct of dogmatism or closed-mindedness was regarded by Allport and Odbert (1936) in their discussion of personality trait names as clearly being a real trait of personality. However, most of the investigation of dogmatism, as an independent variable in its own right, dates from Rokeach's (1954) definition and development of the Dogmatism Scale. Rokeach describes dogmatism as a hypothetical cognitive state which mediates objective reality within the individual. He defined dogmatism as "(a) a relatively closed cognitive organization of beliefs and disbeliefs about reality, (b) organized around a central set of beliefs about absolute authority which, in turn, (c) provides a framework for patterns of intolerance and qualified tolerance toward others" (p.195). He distinguishes dogmatism from rigidity on the one hand, and intellectual conviction on the other. His definition also indicates his view that a closed belief system is the cognitive basis for authoritarianism and intolerance. Weir (1960) argues that the open mind greatly assists learning. "To learn, the individual must be open to new experience .... He must be able to acknowledge new ideas and to give them fair and objective consideration, even though they challenge his deepest convictions" (p.429).

Ehrlich (1961), testing the hypotheses that dogmatism would be inversely related to the degree of learning in a classroom situation, and that such a relationship would be independent of academic aptitude, found both his hypotheses confirmed. Fillenbaum and Jackson (1961) found that students low on dogmatism perform more efficiently on a problem-solving task than students high on dogmatism. Kemp (1960) found a similar relationship between dogmatism and critical thinking. Mouw (1969) found interaction between dogmatism and the level of cognitive process. Open-minded students, as compared with closed-minded students, were found to increase in their ability to perform cognitive tasks as the tasks became more autonomous.

Introversion/Extroversion

One of the notable factor analytic approaches to the study of personality has been that of H.J. Eysenck. His factor analysis of data collected in a study of 700 neurotic soldiers gave rise to two factors described as introversion/extraversion and neuroticism. (The less idiosyncratic spelling of extroversion will be used in the following text.) Later studies refined these factors and Eysenck (1967) argues that there is a genetic basis for both factors. Although Eysenck and Eysenck (1969) have moved away from a two-factor structure by subdividing extroversion into sociability and impulsivity and by adding a dimension of psychoticism (see also Sparrow & Ross, 1964), the studies mentioned below have used instruments based on the original two-factor structure.

Descriptions of the extrovert and introvert personality type can be found in Eysenck and Eysenck (1963) and Eysenck (1965). In summary, the extrovert is described as sociable, impulsive, carefree, and aggressive, and the introvert as bookish, well-ordered, serious, and responsible. These characterizations would seem to suggest that in formal educational settings introverts should be better learners than extroverts. Among studies which support this suggestion are those of Franks (1957) in his laboratory studies on conditioning, and studies made of university students by Furneaux (1956). Savage (1962), in reviewing the American literature on personality factors and academic performance, also concludes that, in higher education, introverts perform better than extroverts. However, a review of the literature by Eysenck and Cookson (1969) suggests that in the late primary school and early secondary school the extrovert tends to perform better. As well as a possible link between neuroticism and poor performance at this age level, Entwistle and Cunningham (1968) found that there were important sex differences in the relationship between introversion/extroversion and attainment. Extroverted girls and introverted boys tend to be more successful in school work than their opposites. However, as indicated below stable extroverted girls and stable introverted boys showed the highest mean attainment scores.

Neuroticism

Lynn and Gordon (1961) and Savage (1962) have both found a curvilinear relationship between neuroticism and achievement for university students, high achievement being associated with an intermediate level of neuroticism. Savage used examination results as the criterion of success, and Lynn and Gordon results on Raven's Progressive Matrices. With school children, however, Entwistle and Cunningham (1968) found that there was a linear and inverse relationship between neuroticism and attainment, a finding which supports those of Savage (1966) and Rushton (1966).

The foregoing discussion indicates that personality characteristics such as anxiety, dogmatism, introversion/extroversion, and neuroticism have been found to affect learning, and the earlier discussion in section 3.4 indicates that they may interact with teaching styles, and so measures of such characteristics should be included in this investigation. What measure or measures should be used?

5.3 MEASURES OF PERSONALITY

Tyler (1969) observes that the measurement of individual differences in personality and motivation has run into more obstacles than the measurement of ability. The most commonly used technique has been the questionnaire. Subjects have been asked to respond to questions about their reactions, attitudes, and feelings towards situations and these responses have been scored in particular ways.

A great many different characteristics have been explored, with protagonists attesting to the conceptual validity and the practical utility of each new dimension. Eysenck (1953) and Cattell (1957) have attempted to bring some order into the chaos by adopting a factor-analytic approach to the isolation of fundamental personality dimensions. Although they differ in the sources of data they have used for their analyses, and in their adoption of orthogonal (Eysenck) or oblique (Cattell) solutions, many writers have remarked on the similarities in a number of the personality constructs they propose. The fact that Cattell's 17 primary source traits are correlated permits further factor analysis, and the extraction of second and higher-order factors. There are clear conceptual relationships between Eysenck's neuroticism, and introversion/extroversion constructs and Cattell's second-order factors I and II, anxiety and extroversion.

The widely used Minnesota Multiphasic Personality Inventory (MMPI) determined its scoring procedures by comparing the responses of deviant groups with those of equivalent normal groups. Factor analysis of correlations between the scales of the MMPI shows that most of the variance can be explained in terms of two major factors. Tyler comments:

> There has been some difference of opinion about what these two factors should be named, but the pattern of factor loadings suggests that Eysenck's characterization of the two major personality dimensions - neuroticism and extraversion - would fit the factors very well. (p.642)

However, "more suitable for research on school populations is the California Psychological Inventory (CPI)" (Tyler, ibid.).

Closely related to the CPI is the Omnibus Personality Inventory (OPI). Taft (1969) refers to them as "blood sisters, 'by California (Berkeley) out of Minnesota'" (p.2). Both the CPI (Gough, 1956) and the OPI (Heist & Yonge, 1962) have been derived in a similar manner to the MMPI. They are both multiphasic personality inventories in which the test items have been derived by empirical means (Heist & Yonge, 1962; Taft, 1969), but whereas the MMPI is primarily intended for use in clinical work, the CPI and the OPI are applicable to normal personalities. In a review of the CPI and the OPI, Taft (1969) observes that they are both intended for measuring personality in the course of guidance and selection work and for research in personality, and that the choice between using one or the other should be related to the purpose in hand. Because in the OPI "almost all dimensions included ... were chosen either for their particular relevance to academic activity or for their general importance in understanding and differentiating among students in an educational context" (Heist & Yonge, 1962, p.1), the OPI was used in the present investigation.

Cronbach and Snow's (1969) review of studies which explore possible interactions between personality characteristics and instructional treatments is prefaced by reservations about conceptualizing personality in terms of traits. They suggest that characteristics such as anxiety may depend to some extent on the specific situation in which the person finds himself. Endler, Hunt, and Rosenstein (1962) have demonstrated, for example, that different persons give evidence of being anxious in different test situations, such as when the test situation poses an academic threat, a social threat, or a physical threat. In completing the OPI, students are asked to respond to items which are set in an academic context and which refer to intellectual, emotional, or social factors which may influence learning.

5.4 THE OMNIBUS PERSONALITY INVENTORY

The OPI, Form F (Heist & Yonge, 1962) consists of 385 true-false items which are scored on 14 scales. Although the OPI manual is vague on the methods by which the original scales were refined, nevertheless there are extensive data on the relationship of the final scales to measures contained in 11 other instruments widely used in personality measurement.

There are three scales on intellectual and scholarly interests: interest in ideas and reflective thought (TI); interest in the use of abstractions and problem solving (TO); and aesthetic interests (Es). And there is a measure of orientation or perceptual style (Co) which is related to a flexible, open approach to phenomena. In order to capitalize on these characteristics it is postulated that an individual must have freedom or readiness to deal with cognitive experiences. A scale concerned with personal autonomy and independence of thought and judgement (Au) is included, and so also is one which, by sampling religious belief and related attitudes, attempts to determine the degree of commitment to fundamentalist thinking (RO).

Several measures of social-emotional adjustment are included that are known to be relevant to scholarly style and behaviour: social extroversion (SE), impulsivity (IE), psychological adjustment (PI), and anxiety level (AL). An additional social attribute that was judged to be relevant was a concern for the well-being of others (Am). No rationale is given for the inclusion of a scale which measures a practical and utilitarian outlook (PO). Nevertheless this scale has high negative correlations with TI, Au, and Co and appears to have a high authoritarian component. The scale sampling masculine-feminine interests and attitudes (MF) correlates highly with the Es scale, largely because there is a substantial overlap of items: of the 24 items on the Es scale, 10 appear on the MF scale.

The validity of self-report testing was considered by Heist and Yonge and they have included a response set scale (RB) which focusses on an individual's need to make a good impression.

Heist and Yonge (1962, p.49) give two estimates of internal consistency for the OPI scales, and two of test-retest stability. On the total standardization sample (N = 7283) the Kuder-Richardson Formula 21 coefficients ranged from 0.67 to 0.89 (Mdn = 0.83), and on a sample of freshmen (N = 400) the corrected split-half correlations ranged from 0.65 to 0.91 (Mdn = 0.83). In the test-retest calculations the reliability coefficients ($N_1 = 67$, $N_2 = 71$) range from 0.84 to 0.93 ($Mdn_1 = 0.87$) and 0.79 to 0.94 ($Mdn_2 = 0.89$) respectively.

Having been developed in the USA, and standardized against a population of college freshmen, there was some need to try out the OPI with 11th grade Australian students to check on any difficulties with unfamiliar language and concepts and to note the time requirements for administering the test. Such a trial was given to an 11th grade biology class in the November preceding the year of the investigation. Students were requested, as soon as they came to an unfamiliar word or idea, to ask the investigator for its meaning which was then written on the chalkboard for other students to see. In the conduct of the final investigation the following direction was given to students:

> This test has been designed for use in American schools and some of the words and phrases may be strange to you. On the blackboard, I have written down the common meanings of the more difficult ones. If you came to a word you don't understand you should check its meaning with the list on the board. (Appendix I)

This direction and the list of words is found in Appendix I. During the administration of the OPI there were no further requests for clarification of the meanings of words and phrases.

The manual (Heist & Yonge, 1962, p.50) reports substantial correlations between a number of the scales. These intercorrelations are of no concern in psychological counselling. In fact Taft (1969), in his review of the CPI and the OPI, remarks that "these inventories have the advantage over factorially based inventories ... in that they are more closely tied to actual behaviour and socially observable qualities, or 'folk concepts' as Gough calls them" (p.5). However, for use as individual predictors in a multiple-regression analysis orthogonal variables are desired. Although there is no requirement in multiple regression that the predictor variables be uncorrelated (Darlington, 1968), measures such as the F value, R^2, and regression weights, and the notion of 'contribution to variance' become ambiguous if they are not. One approach is to take advantage of the intercorrelations between the scales and by factor analysis to reduce them to fewer dimensions. However, Thurstone (1947) points out that correlations used in factor analysis should be between variables that are linearly independent. An examination of the 14 scales of the OPI show that, of the 385 items, 163 appear in more than one scale. Guilford (1952) observes that the contribution made to the obtained correlation between two scales is legitimate if the item measures factors common to both scales, but that the item's contribution to any total score also includes some specific and error variance which would inflate the value of the correlation obtained. Considering these objections the item overlap between the 14 scales may be considered excessive.

An examination of the scales reveals that the first 10 scales comprise 354 of the total of 385 items but contain only 21 items that appear in more than one of these scales. In the OPI Manual the first six scales (TI, TO, Es, Co, Au and RO) are postulated as important measures of characteristics associated with intellectual functioning and have been used to generate a set of Intellectual Disposition Categories. The next four scales (SE, IE, PI and AL) all appear to measure personal characteristics whose relationships with academic performance have been suggested by the research discussed in section 5.2. SE and PI are described in terms similar to Eysenck's (1965) dimensions of extroversion and neuroticism and anxiety is measured here by scale AL. The work of Sanford, Webster, and Freeman (1957) suggested the inclusion of a measure of 'impulsivity' in the OPI. This consideration received additional support from the work of Eysenck and Eysenck (1969) quoted above. The OPI Manual proposes "a measure of impulsivity ... as being essential in the assessment of 'psychic energy' and some of the bases for creativity, as well as for its possible relationship to the adaptation or adjustment demanded by some of the restrictive, over-controlled learning environments" (p.3). The remaining four scales (Am, PO, MF and RB) are composed substantially of items which already appear in the preceding 10 scales. Am scale has 36 items, 21 of which appear in the preceding 10 scales, PO (30, 23), MF (56, 50) and RB (28, 27).

In these circumstances it was decided not to use scores on these latter four scales but to carry out a principal components factor analysis of scores of the 146 grade 11 biology students on the first 10 scales of the OPI (TI, TO, Es, Co, Au, RO, SE, IE, PI and AL).

The first four roots extracted 72.6% of the trace. The varimax factor loadings based on the correlations between scores on the 10 OPI scales are shown in Table 5.4.1. In addition to the communality for each variable (scale), an index of factorial simplicity (Kaiser, 1974) is given. This index varies between zero and one and reflects the tendency towards unifactoriality, that is, towards each row of the factor matrix having only one non-zero loading. Although much the same interpretation may be made by inspection of the factor matrix, this index acts as additional support in interpreting less simple matrices than the one presented below. Kaiser's alliterative rule of thumb for evaluating the levels of the index of factorial simplicity merits quotation. He suggests;

> in the .90s, marvelous
> in the .80s, meritorious
> in the .70s, middling
> in the .60s, mediocre
> in the .50s, miserable
> below .50, unacceptable. (p.35)

TABLE 5.4.1

LOADINGS ON FACTORS EXTRACTED FROM TEN OPI SCALES

OPI Scale		Factor Loadings				Communality	Index of Factorial Simplicity
		1	2	3	4		
1	TI	0.10	-0.12	-0.89*	0.10	0.83	0.94
2	TO	0.07	-0.16	-0.73*	-0.18	0.59	0.86
3	Es	-0.02	0.15	-0.61	0.53	0.67	0.55
4	Co	0.65	0.05	-0.43	-0.05	0.61	0.65
5	Au	0.62	-0.15	-0.17	-0.45	0.64	0.54
6	RO	0.11	-0.07	0.03	-0.78*	0.63	0.96
7	SE	0.49	-0.47	0.10	0.58	0.81	0.34
8	IE	0.82*	0.23	0.12	0.07	0.75	0.87
9	PI	-0.11	-0.92*	-0.02	-0.02	0.86	0.98
10	AL	-0.02	-0.92*	-0.13	-0.05	0.86	0.97

Note: Salients are marked with an asterisk and highest loadings are underlined.

Tentative descriptions of these orthogonal factors are given below, based on the descriptions of responses to items in the OPI scales in the OPI Form F Manual. Their short titles are aids to discussion, and are not meant to imply an identity with constructs of the same names from other instruments.

Factor 1 (Impulsivity)

OPI scale	Factor loading	Index of factorial simplicity
Impulse expression	0.82	0.87
Complexity	0.65	0.65
Autonomy	0.62	0.54

This factor can be described principally in terms of Impulse Expression. High scorers admit to expressing their impulses readily, a characteristic which Yonge and Heist (1962) suggest as being among the bases of creativity. The contributions of Complexity and Autonomy suggest a tendency to an experimental and flexible orientation and an independent viewpoint.

Factor 2 (Neuroticism)

OPI scale	Factor loading	Index of factorial simplicity
Personal Integration	-0.92	0.98
Anxiety Level	-0.92	0.97

High scorers on this factor (low scorers on the contributing OPI scales) often feel inadequate and are sometimes debilitated by self doubts. They are often nervous and worried and take rebuffs hard. The lesser contribution of SE also suggests that they tend to withdraw from social contacts and responsibilities. A low score on this factor represents emotional maturity, good psychological adjustment and low anxiety.

Factor 3 (Non-intellectualism)

OPI scale	Factor loading	Index of factorial simplicity
Thinking Introversion	-0.89	0.94
Theoretical Orientation	-0.73	0.86
Estheticism	-0.61	0.55

High scorers on this factor (low scorers on the contributing OPI scales) dislike serious thought or involvement in philosophical or scientific issues. The loading of Estheticism also suggests a lack of interest in literary-aesthetic pursuits. Persons obtaining low scores present a reflective and theoretical orientation to the world of ideas and scholarship.

Factor 4 (Dogmatism)

OPI scale	Factor loading	Index of factorial simplicity
Religious Orientation	-0.78	0.96
Social Extroversion	0.58	0.34

This factor can be described principally in terms of the characteristics of low scorers on the Religious Orientation scale. Social Extroversion can be seen from its index of factorial simplicity to be a complex variable with moderate loadings on the three non-intellectual factors. High scorers on this factor tend to be dogmatic, conservative, and orthodox religious believers. The loading of SE suggests that they also tend to be gregarious. Low scores represent a tendency to be sceptical of fundamentalist beliefs, and independent of traditional authority, and a preference for being and working alone.

5.5 COGNITIVE STYLE: COGNITIVE PREFERENCE

The third class of personological variables mentioned by Tyler (1969) is that of cognitive styles. There appears to be a number of terms such as learning style, perceptual attitude, and cognitive control that are used more or less synonymously with cognitive style, and the term cognitive style itself has different meanings for different authors.

Witkin, Dyk, Faterson, Goodenough, and Karp (1962) whose research began with the isolation of a perceptual trait, field-dependence/independence, came to believe that perceptual characteristics were just one aspect of an individual's level of 'psychological differentiation'. Witkin, Moore, Goodenough, and Cox (1977) comment:

> Extensive evidence, accumulated over the years, shows that the styles we first identified in perception, manifest themselves as well when the person is dealing with symbolic representation, as in thinking and problem solving. The individual, who, in perception ... is relatively field dependent ... is likely to have difficulty with that class of problems ... where the solution depends on taking some critical element out of the context in which it is presented and restructuring the problem material so that the item is now used in a different context. (p.8)

Because the field-dependent/independent characteristic encompasses perceptual and intellectual functioning they refer to it as a cognitive style. Holzman and Gardner (1960) refer to the concept of 'cognitive control' as being based

> on the observation that individual differences in cognitive behaviors often reflect habitual, but not necessarily conscious strategies of adaptation, and the further observation ... that the enduring response dispositions represented by such strategies serve to mediate the effects of drives upon cognitive behavior. (p.176)

Ausubel (1963) refers to cognitive style as "the self-consistent and enduring individual differences in cognitive organization and functioning" (p.95). Kagan, Moss, and Sigel (1963) have used the term cognitive style to refer to stable styles of conceptualizing the environment in problem-solving situations, analytic and non-analytic. In his later writing, Kagan (1965) describes the influence of the reflection/impulsivity dimension in problem-solving, a dimension which has been referred to earlier in this chapter as a personality variable. Messick (1970) summarizes, from the research literature, some nine dimensions of individual differences that appear to reflect consistent differences in the mode or manner of cognition as distinct from the level of skill displayed in the cognitive task, and concludes that "cognitive styles, for the most part, are information-processing habits. They are characteristic modes of operation which, although not necessarily completely independent of content, tend to function across a variety of content areas" (p.190). After more detailed analysis of one of these dimensions (Witkin's 'articulated' versus 'global' cognitive style) and its pervasiveness in the research literature in personality and psycho-pathology, he concludes that

> cognitive styles, by embracing both perceptual and intellectual domains and by their frequent implication in personality and social functioning, promise to provide a more complete

> and effective characterization of the student than could be obtained from intellectual tests alone. (p.195)

One aspect of cognitive style of particular relevance to learning in schools is Heath's (1964) construct of cognitive preferences: those "differing modes of attending to the subject matter of the course" (p.241). Cognitive preferences are also expected to reflect differences in students' cognitive structures.

One important issue in the consideration of cognitive style as a variable in educational research is whether or not the characteristic measured is stable over time. Witkin, Goodenough, and Karp (1967) in their studies of psychological differentiation establish that there is a progressive decrease in field dependence up to age 17 which then levels off through young adulthood. However, the characteristic is stable in the sense that each individual tends to maintain his own position relative to his peer group from age to age: "level of differentiation, as assessed in perceptual functioning, shows high stability during the growth years, even over a time span as long as fourteen years" (p.298). Siegel and Siegel (1965) in their examination of cognitive style, conclude that by the time a student is an undergraduate "his educational set can be assessed reliably indicating that this is a stable rather than a transitory variable" (p.11). Williams (1975) also finds evidence for the stability of cognitive preferences.

Yando and Kagan (cited in Glaser, 1972) found, however, that first-grade children placed with teachers who have a reflective style become more reflective than children placed with impulsive teachers. In comparing the cognitive preferences of students after taking PSSC physics courses with those of a control group after taking non-PSSC courses, Heath (1964) found that the students taking the PSSC courses had developed significantly different cognitive preferences. Marks (1967) found a similar result in chemistry. Scott (1973) also provides evidence that cognitive style can be influenced by an inquiry-centred teaching strategy. Mackay (1971b) found that over a two-year period, "higher aptitude pupils are more likely than low aptitude pupils to modify their cognitive styles so as to increase the degree of congruence between their own cognitive styles and those emphasized during instruction" (p.13).

Such changes in cognitive preferences, however, may not be transferable: Atwood (1967b) could detect no shifts in cognitive preferences in chemistry during a one year period in which students were studying physics. And, classroom variables do not always modify cognitive preferences in the same subject: Schmedemann and La Shier (1967) found no significant relationships between students' perceptions of their teachers' classroom behaviour and their cognitive preferences in that subject.

The evidence on the stability or modifiability of cognitive styles, then, is not unequivocal, and if there is a suggestion that training can modify these personological variables they may not be very different in this respect from intelligence and other aptitudes, which tend to be stable unless subject to consistent and prolonged attempts to promote change. The main theme of Glaser's (1972) article is "the importance of modifiable behavioral processes in adaptive education as opposed to notions of relatively fixed intelligence and aptitude" (p.10). A survey carried out by Glaser and Resnick (cited in Glaser, 1972) indicates that there is "a significant shift, as compared with a few years ago, in the balance of evidence concerning the trainability of these [basic aptitudes and Piagetian] processes" (p.10). Even if Glaser's point is accepted, that these basic aptitudes and processes are modifiable, many authors have no doubt that they are important personological variables influencing new learning. For example, Ausubel (1964) confidently asserts that a learner's cognitive structure is "the most significant independent variable influencing the learner's capacity for acquiring new knowledge" (p.237).

There are differences in the definitions of cognitive style adopted by the various authors, and differences in aspects of cognitive style studies. These differences make it difficult to compare the various studies and to generalize data or conclusions from them. However, the accumulation of research in the area, reviewed in recent studies by Mackay (1971b), Koay (1975), and McNaught (1975) adds to the conviction that cognitive style is an important class of individual differences which is related to achievement, and which should be sampled in this study. For example, Witkin et al. (1977) suggest that "students may learn more effectively when taught by teachers matched to them in cognitive style" (p.34). Barnett (1974) comments: "students who show strong preference for questioning and application could be expected to feel more comfortable, and to work more effectively using inquiry, than students who show a strong preference for memory" (p.141). In their study of cognitive preferences in science, Walters and Sieben (1974) observe that "it is becoming more evident that not all children learn science the same way. The 'free and unguided exploration' approach may well be appropriate for some students, but may be inappropriate for others" (p.73).

The literature provides support for the contentions that aspects of cognitive style can be identified experimentally, that these aspects are enduring dispositions that are relatively stable over time unless subject to instructional procedures designed to produce change, that they are important independent variables influencing new learning, and that they may have greater interactive effects with instructional treatment than measures of conventional aptitudes such as intelligence.

Heath's (1964) work on cognitive preferences grew out of a consideration of the objectives of the new science curricula which developed in the United States of America in the post-Sputnik years. This new generation of science curricula in America has its parallel in the Nuffield Science and Mathematics Project in Britain, and science curriculum development in Australia has been greatly influenced by both the American and British experience. Among the more obvious features of the new courses is the emphasis placed on the nature of scientific inquiry. Students are required to use genuine problem-solving procedures in the laboratory rather than merely following a 'cook-book' style of laboratory investigation. There has also been a tendency to reduce the range and detail of the substantive content of courses in favour of a greater emphasis on the major organizing principles of the science, an emphasis cogently argued in the study of biology by Bonner (1963) and Novak (1963). This emphasis also brings into relief the structure of knowledge, an aspect of curriculum that has received considerable attention since the publication of The Process of Education (Bruner, 1960). It is claimed by the proponents of the new courses, and by authors such as Schwab (1962), Ausubel (1965), Bruner (1966), Heath (1964), Suchman (1964), Scott (1964, 1973), and Atwood (1967a), that the new courses call for different modes of cognitive functioning in the acquisition, processing, and transmission of information; that is, the courses favour particular cognitive styles.

Heath (1964) observes that "the cognitive preferences exhibited by students in the conventional and the new science curriculums may be expected to differ significantly" (p.241). Heath was not interested so much in whether a student could identify correct information or not, but rather in the manner in which he was likely to deal with the information intellectually. He selected four classes of cognitive preferences relevant to the study of secondary school science courses:

> Preference M (Memory): Preference for remembering a specific fact or term related to information provided, without considering its implications, application, or limitations.
>
> Preference A (Application): Preference for identifying a practical application related to the information provided in a general social or scientific context.
>
> Preference Q (Questioning): Preference for critically questioning or challenging the information provided, or recognition of a limitation or extension of the information.
>
> Preference P (Principles): Preference for identifying the fundamental principle underlying the information provided, or for drawing conclusions based on the underlying principle.

Heath's original Cognitive Preference Test: High School Physics consisted of 20 items. Each item presented introductory information in the stem, sometimes illustrated with a diagram or graph, which was followed by four options. However, each of the four options which followed was a correct statement and the student, knowing this, was asked to choose the option he preferred most in connection with the introductory statement. In each item there was one option that was designed to appeal to students showing each of the cognitive preferences outlined above, and a student's score on each cognitive preference was merely the number of items in which that particular preference had been indicated. In the study which used this instrument Heath (1964) found that "this type of instrument can identify, in meaningful context, curriculum-related differences in cognitive style" (p.250). Tamir (1975) sees cognitive preferences as most useful variables in assessing the outcomes of a specific curriculum in that they can be related to instructional treatments more readily than variables measured by more general tests such as the Watson-Glaser Critical Thinking Appraisal. These cognitive preferences have been used in a number of studies in different fields of science: general science (Davids, 1968; Ogunyemi, 1973; Walters & Sieben, 1974); physics (Heath, 1964; Schmedemann & La Shier, 1967; Atwood, 1969; Mackay, 1971b, 1972); chemistry (Atwood, 1967b; Marks, 1967; Mackay, 1972; Kempa & Dubé, 1973; McNaught, 1975); and biology (Barnett, 1974; Tamir, 1975; Mackay, Note 6).

Mackay (1971b) examined three methods of responding to Heath's cognitive preference test: the selection of the most appealing response, the selection of the most appealing and least appealing responses, and rank-ordering all four options. On the basis of his investigations he suggests that students be asked to select the most appealing and the least appealing response to each item. The score on each preference is then calculated as (the number of items) + (the number of items on which the option for that cognitive preference is selected as the most appealing) - (the number of items on which the option for that cognitive preference is selected as the least appealing). This yields a score on each cognitive preference between values of 0 and twice the number of items.

The four scores obtained from Heath's cognitive preference test reflect a student's relative preferences for the four classes of cognitive preferences rather than an absolute scaling for each preference, since in any one item one choice precludes any other. There are problems in the statistical treatment of such ipsative scores that are not present when scores belong to independent normative scales. Wish (1964) and Mackay (1971b) discarded the magnitudes of the scores and used instead a rank order of preferences for the individual. Mackay then sought a two-dimensional solution to the four ranks by an unfolding analysis originally developed by Coombs (1958, 1964) and extended by a number of other authors.

> The best solution derived from this analysis for the location of the four cognitive preferences in two-dimensional perceptual space corresponded to the cognitive preference stimuli for Applications (A) and Questioning (Q) being located along a dimension which was approximately orthogonal to the dimension along which the cognitive preference for Principles (P) and Memory (M) were located. This solution could accomodate all 285 rank orderings of class means on the Cognitive Preference Test Physics and the rank orderings of 92.4% of the 9263 individuals involved in this study. The results of two Q-technique factor analyses conducted for sub-samples of this group yielded an identical solution. (Mackay, Note 6, pp.21-22)

The appropriateness of cognitive preferences being represented on these two dimensions has also been confirmed in chemistry classes by Kempa and Dubé (1973) and Rogel (cited in Tamir, 1975), and in biology classes by Mackay (Note 6), although Tamir's (1975) two-dimensional solution is somewhat different.

In this study, the Test of Cognitive Preferences (TOCP) Biology developed by Mackay (Note 6) was used. To produce a score on Scale Q - A, the separate scores for Q and A are calculated as above, the A score is subtracted from the Q score, and a constant of 20 is added to ensure positive values. The score for Scale P - M is similarly calculated as (P - M + 20). This procedure yields scores on Scale Q - A and Scale P - M with values between 0 and 40.

> A score on Dimension 1 (Scale Q - A) can be interpreted as representing a respondent's relative cognitive preference for theoretical as compared to applied presentations of scientific material, or his preference for a scientific as compared to an engineering orientation. A score on Dimension 2 (Scale P - M) can be interpreted as representing a respondent's relative cognitive preference of scientific phenomena in terms of the fundamental scientific principle involved, as compared to rote memory tasks. (Mackay, Note 6, p.31)

Estimates of the reliabilities of the two scales, Q - A and P - M, are given in Table 5.5.1.

TABLE 5.5.1

ESTIMATES OF RELIABILITY OF TOCP BIOLOGY SCALES

Sample	N	Cronbach α	
		Scale P - M	Scale Q - A
Mackay (Note 6)	735	0.61	0.38
Present investigation	146	0.63	0.65

Cronbach's α reliability coefficient (Cronbach, 1951) gives the same values as the KR-20 (Kuder & Richardson, 1937) where items are scored right/wrong, but it can also be applied where, as with the TOCP, other scoring procedures are adopted. It also has an advantage over the split-half coefficient in that it can be shown to be equal to the average of all split-half coefficients calculated from all possible splits of a test, and so does not depend on the choice of split (Lord, 1956). These reliabilities are somewhat low and may be due, in a large measure, to the comparative brevity of the tests: each scale is contributed to by only 20 responses. Tamir (Note 7) has expressed his reservations about the length of the TOCP. He reports Cronbach α reliability coefficients for his 40-item Biology Cognitive Preference Test as 0.82, 0.80, 0.84, and 0.73 for the scales of Recall, Principles, Questioning, and Application respectively (1975, p.240). However, this longer test was not available when the field work for the present investigation was being carried out.

CHAPTER 6

INSTRUCTIONAL TREATMENTS

6.1 TREATMENT VARIABLES: CLASSROOMS, COURSES, AND TEACHERS

6.2 TEACHER VARIABLES: STRATEGY AND STYLE

6.3 CONTRASTING STYLES

6.4 OBSERVING CLASSROOM BEHAVIOUR

6.5 SOME CATEGORY SYSTEMS

6.6 THE DEVELOPMENT OF A CATEGORY SYSTEM

6.7 RECORDING AND CODING PROCEDURES

6.8 VALIDITY AND RELIABILITY

6.9 VALIDATION AND DESCRIPTION OF TREATMENTS

6.1 TREATMENT VARIABLES: CLASSROOMS, COURSES, AND TEACHERS

Siegel (1960) delineates four major classes of variables which should be considered in experimental research on teaching: classroom environments, student variables, instructor variables, and course variables. Student variables have been considered in some detail in the previous chapter; the three remaining variables, which can all be considered as aspects of the treatment, will now be discussed.

The physical environments of all classrooms in the study were very similar. The biology classrooms used by six classes were built for the Victorian Education Department, during the late 1950s and early 1960s, to a standard architectural plan and so are virtually identical in terms of built-in physical provisions. The remaining government school and the non-government school were housed in older buildings, but had closely similar provisions since they had met the standards for accreditation in order to enable them to present students for the Leaving Certificate by internal rather than public examination. The provision of laboratory equipment and teaching equipment, such as microprojectors, in government schools is largely standard and in the other school is similar being governed by the requirements of the course prescriptions. The laboratory manuals and Teacher's Guides give detailed lists of apparatus and materials that are required for the practical exercises. Most classes contained one or two repeat students and some who were not present for the full year. The sizes of the classes used in the research (with the additional repeat students, and those for whom complete data were not available, given in brackets) were: 19(2,4); 15(1,0); 15(1,2); 14(2,3); 20(2,2); 11(1,3); 25(0,0); and 27(2,1).

The introduction and development of the Web of Life biology course in Victoria made it possible to obtain classes for the present research with practically no variation in course content and objectives.

From 1967 to 1972 the schools followed closely the Web of Life course prescribed for both 11th and 12th grades by the Victorian Universities and Schools Examinations Board. Some latitude in the selection of content was encouraged in 1973, so it was necessary to select classes whose teachers were intending to follow the original prescriptions. The content of the course is set out in detail in Biological Science: The Web of Life, second edition (Morgan, 1973); in the associated Student's Manual, part 1 and 2 (Australian Academy of Science, 1969); and the Teacher's Guide, parts 1 and 2 (AAS, 1969-1976). The objectives of the course are described in the Biology Teachers' Handbook (Klinckmann, 1970, pp.9-53) and in specific behavioural terms in the Teacher's Guide, parts 1 and 2.

6.2 TEACHER VARIABLES: STRATEGY AND STYLE

In terms of classroom environments, and course content and objectives, the biology classes investigated in this research were very similar. The greatest differences lay in the teachers. Mitzel (1960) makes a distinction between presage and process variables. Characteristics which exist before the teaching begins are presage variables whereas classroom behaviours are process variables. While recognizing the importance of presage variables such as the teacher's personality, intelligence, training, experience, and cultural and socio-economic background, this research is fundamentally concerned with process variables, the teachers' classroom behaviour. In addition to the specification of course content and objectives, considerable advice is given in the curriculum materials on general teaching strategies to promote the inquiry-centred objectives of the course. See particularly the Teacher's Guide, part 1 (AAS, 1973, pp.6-23), the Teacher's Guide, part 2 (AAS, 1976, pp.9-65), the Biology Teachers' Handbook (Klinckmann, 1970, pp.330-394) and sections of the textbook, Student's Manuals, and Teachers' Guides relating to the selected content. During the years 1965-67 in-service courses for practising teachers of biology were held to introduce them to the rationale, approach, and content of the new course, and many sessions were devoted to the teaching strategies that were considered necessary to achieve the inquiry-centred objectives. Teachers trained in colleges and universities since that time have had their biology curriculum studies centred on the Web of Life materials.

The use of the phrase 'teaching strategies' rather than 'teaching methods' is intended to emphasize, in the manner of Taba and Elzey (1964), the deliberate selection and organization of teaching techniques that are designed to achieve specified outcomes in terms of student behaviour. For example, questioning is a teaching technique or method that is used by nearly all teachers. However, it is clear that some questions merely require a student to recall what he has learned, whereas other questions may require him to interpret this material, to apply it in new situations, to analyse it, to organize it in new ways, to evaluate it. Similarly, some questions are asked in the knowledge that there is a correct answer whereas other questions may have several acceptable answers, or an answer may not be known. The deliberate choice of a particular kind of questioning in an attempt to have students think in a particular way would be regarded as an adoption of a specific teaching strategy. Even where teachers have consciously adopted the same content and objectives of a course and the accompanying teaching strategies that are described in the curriculum package, their classroom behaviours may still show considerable variation (Yager, 1968). These remaining differences in teaching behaviour are referred to in this investigation as 'teaching style', and were the focus of attention during the development of a category system for recording and analysing teaching behaviour, described below. The distinction between teaching strategies and teaching style is of some importance because the curriculum materials used in this investigation suggest specific strategies, particularly in the conduct of laboratory investigations which are the bases of the course. The Teacher's Guide, part 2 (AAS, 1976), after giving advice on teaching strategies says:

> Teaching style differs from person to person and there may be as many teaching styles as there are teachers. But precisely what one does in a given situation requires a continual effort to adapt one's own personality, style and background to the requirements of the teaching role. (p.33)

Klinckmann (1970) comments:

> It is much more difficult to identify and describe teaching styles than teaching strategies, since the former are more probably closely related to teachers' personalities. (p.334)

In other studies where the content, objectives, and teaching strategies have not been so specifically determined the term teaching style has generally had a wider meaning.

Teaching style is not an easy concept to define and a number of varying descriptions are found. For example, Oliver and Shaver (1962) define two contrasting styles of teaching, recitation teaching and Socratic teaching, in terms of statement posture, discussion posture, statement types, intellectual operations, the procedural dimension, and the affective dimension. Their descriptions of Socratic and recitation teaching would be referred to as strategies by Taba and Elzey (1964) and Suchman (1960), and as 'patterns of moves' by Coombs (1966). Turner (1971) distinguishes between task-relevant style, which concerns variations in teaching operations such as questioning or evaluating responses, and leadership style, which concerns variations in the way a teacher operates with a group, such as being subject-centred or student-centred, indirect or direct. He points out that in programmed instruction there are only variations in task-relevant style, but that the social context makes teaching an instance of leadership in task-oriented groups. Turner's distinction becomes, basically, not that between task-relevant and leadership styles, but between on-task leadership and off-task leadership. Shumsky (1968), on the other hand, has described three contrasting styles as follows. The first is proposed as a stereotype to illustrate extreme concentration on specificity and recall with an emphasis on a high degree of structuring by the teacher, reliance on the text, and on a fixed-answer problem solving approach. The second style emphasizes pupil discovery and concern with the process of inquiry, and moves quickly from the data to a consideration of constructs and relations. The third style emphasizes personal interpretations, independent and divergent thinking, and pupil involvement at all stages of planning and executing a project. These three styles can be referred to as authority-centred, inquiry-centred,and creative styles.

These formulations illustrate the difficulties of separating teaching approaches or strategies from the rather more idiosyncratic or individual behaviours of teachers, which can be referred to as their teaching style. Yager (1968) and Thomas and Snider (1969) provide some evidence that even after the conscious adoption of particular teaching strategies as defined by Taba and Elzey (1964) there still remains a pattern of behaviour characteristic of the individual teacher that is difficult to modify. In a day-by-day prescription of content, emphasis, and activities, Thomas and Snider (1969) set out to compare the results of 'guided discovery' versus 'didactic' methods of teaching. Two teachers used both methods, but despite the prescriptions, they found such differences between the teachers, as measured by student gain in factual-conceptual achievement, that they did not attempt to combine the data for each teaching strategy. Yager (1968) used a modified Flanders (1965) instrument to identify the styles of teachers who had agreed to the same objectives and were using the same materials.

In the present research it is held that, after assent has been given by the participating teachers to adopt the inquiry-centred teaching strategies that pervade the curriculum materials of the Web of Life biology course, there remains in each teacher habitual patterns of inter-personal reaction and classroom technique that are characteristic of that teacher. These habitual patterns of teaching behaviour are referred to as teaching style in the manner of Klinckmann (1970); this is a much more restrictive concept of style than that recently popularized by Bennett (1976).

6.3 CONTRASTING STYLES

As argued in chapter 4, this investigation is committed to the principle of non-manipulation of, and minimal interference with, the natural classroom environment. Therefore in this study it was decided that a comparison of teaching styles could best be accomplished by observing teachers following their own natural styles in a normal school situation over a year, and subsequently placing them in different categories.

In searching for the elusive significant interactions between student characteristics and instructional treatments it seemed best to heed Cronbach's (1967) advice and select sharply contrasting treatments. Consonant with the general purposes of the study, which was concerned with an investigation of aspects of the inquiry-centred objectives of the Web of Life biology curriculum, general descriptions of two contrasting teaching prototypes were generated. The prototypes were developed from the descriptions of the authority-centred and inquiry-centred styles of Shumsky (1968), with reference to Bingman's (1969) work on inquiry objectives in the teaching of biology.

TABLE 6.3.1

PROTOTYPES FOR SELECTION OF TEACHERS

Prototype A	Prototype B
A teacher who	A teacher who
1. is recognized as a good teacher;	1. is recognized as a good teacher;
2. structures the learning situation closely;	2. uses a flexible lesson structure;
3. follows text and laboratory manuals faithfully;	3. uses many materials and sources additional to text and laboratory manuals;
4. places great emphasis on the acquisition of the knowledge and generalizations of biology;	4. places great emphasis on the acquisition of the skills of inquiry;
5. organizes laboratory sessions so that most students arrive at satisfactory conclusions;	5. tolerates a high degree of uncertainty in the results of laboratory investigations;
6. does most teaching with the class as one group;	6. does most teaching by assisting individuals and small groups;
and whose students are well disciplined and quiet.	and whose students are active in discussion and problem-solving.

These lists may appear paradoxical in that it has been said that all participating teachers had given assent to the objectives of the Web of Life course, and to the inquiry-based teaching strategies that were recommended as leading to these objectives. Prototype A does not appear the type of teacher who would do justice to these inquiry-based strategies. This may indeed be so, and illustrates the difference between a conscious choice of strategy and perhaps an intellectual commitment to the importance of the process of inquiry, and more deep-seated and habitual responses which may be related to personality, cultural and socio-economic background, experience, and teacher training. Strategy and style may be consonant, on the other hand they may be dissonant. The research of Yager (1968) and Thomas and Snider (1969) reflects dissonance, or the pervasive and dominant influence of style over strategy. Gallagher (1967), working with six biology teachers using BSCS materials, also found that an individual teacher tends to maintain the same style of teaching in the face of considerable variation in the nature of content and the objectives of different topics.

Two members of the Board of Inspectors of Secondary Schools and the Biology Method Lecturers of the five secondary-teacher training institutions in Victoria were asked to supply a list of names of biology teachers who they believed were appropriately described by Prototype A. Three weeks later they were asked to supply a list of names of biology teachers who they believed were appropriately described by Prototype B. Both prototypes were described in positive terms, and it was hoped that the presentation of Prototype A first, and without reference to Prototype B, would avoid invidious comparisons and assumptions about the value position of the investigator.

All consultants would not necessarily know the same teachers, but when three or more consultants independently included the same teacher on the one list this was taken to be some evidence of consensus.

Eleven such teachers were found who taught in the eastern districts of Melbourne and who were intending to follow the 1972 prescriptions of the Web of Life biology course. Approaches were made to 10 of these teachers, 5 from each list, to secure their consent, and that of their school, to taking part in the project the following year. The class of the 11th teacher was used to time the administration of the pretests, to check on the appropriateness of the American language used in the OPI, and to rehearse the method of observing and recording class lessons with a minimum of disturbance to the class.

After the administration of the pretest at the beginning of the next school year, one class was found to contain a high proportion of recent migrants whose English was very poor. The research sample was reduced to eight classes, each containing both boys and girls, four taught by men and four by women; four teachers coming from each list.

It would now be possible to accept that these two groups of teachers constitute contrasting treatments, and to proceed to test various hypotheses about the effects of authority-centred and inquiry-centred teaching styles on a number of dependent variables. However, the research reviewed in chapter 2, on contrasting instructional treatments, suggests strongly that it is important to substantiate that the contrasting teaching styles are demonstrably different between groups but similar within groups. In addition, it is argued in chapter 4 that in order to make generalizations that are ecologically valid it is necessary to be in a position to provide detailed descriptions of the treatments. In order to meet both these requirements it was decided to make a systematic observation of the teachers' classroom behaviour over the year.

6.4 OBSERVING CLASSROOM BEHAVIOUR

The two principal methods of systematically observing and recording classroom behaviour are by use of 'rating systems' or 'category systems'. Rating systems are regarded as high-inference measures in that they typically focus upon such behavioural constructs as 'enthusiasm', 'clarity of presentation', and 'warmth', and the observer is required to infer the presence of these constructs from behaviour that is shown in the classroom. Category systems, because they focus on classroom behaviours, are regarded as low-inference measures, although in practice most category systems need to specify additional ground rules so that unclear behaviours can be categorized consistently.

Although Rosenshine (1970) has referred to some promising and robust constructs such as 'clarity of presentation', and to the advantage of having considered clues from a variety of sources before making judgements, rating scales suffer from a number of disadvantages which affect their reliability. There is often a lack of clarity about the trait being rated which adds to the difficulty of inferring its presence and strength. There is general difficulty in securing common agreement about qualitative categories such as superior, excellent, satisfactory, poor, unacceptable: raters often react by being cautious and giving a high proportion of central ratings, or by being generous and giving a high proportion of favourable ratings. And there is the well-documented halo-effect, where the general impression gained of an individual leads the rater to award consistently high ratings, or low ratings. Rating involves not only an inference as to whether a particular phenomenon is present or not, but also some judgement as to its quality. Such judgements are best made as the events are taking place, taking account, as

Rosenshine says, of clues from a variety of sources. Such judgements of quality are difficult to make from audiotape, or even videotape recordings. Also in order to provide any estimates of reliability of the ratings two or more raters would need to be in the classroom at the one time. Quite apart from the effect additional observers may have had on the normal classroom behaviour, additional personnel were not available to the investigator. For these reasons it was decided that the use of rating systems was both inadvisable and impractical.

On the other hand, it was possible for the lone investigator to make good quality audiotape records of the teachers' verbal behaviour and, simultaneously, to make a written description of the teachers' non-verbal behaviour. Examples of the teachers' important non-verbal behaviour include writing on the blackboard, assisting silently with laboratory apparatus such as adjusting a microscope or slide, walking around the room in general supervision, smiling encouragement at a student, and so on. Written transcripts could then be made of the audiotapes and annotated with the observer's written record. The combined record could then be examined as frequently as desired and categorised many times and by many persons. It was decided to use such a recording technique combined with a category system for the analysis of the contrasting teaching styles.

6.5 SOME CATEGORY SYSTEMS

The selection of an appropriate system for the analysis of teaching style is not easy. Nuthall and Church (1972) point out that it is a mistake to treat observation systems like published mental tests.

> They cannot be subjected to evaluation on the basis of accepted a priori criteria. We do not know enough about the causes and effects of teacher behaviours to be able to lay down the standards which observation systems must meet

> in order to be considered valid and reliable instruments. (p.504)

This is as true of frequently used systems such as FIAC (Flanders, 1970) and OScAR (Medley & Mitzel, 1963) as of systems developed for a single research study designed by an individual researcher. Nuthall and Church continue:

> The research worker who is interested in studying teaching behaviour must treat any published observation system with caution. In the present state of our knowledge, he must begin by trying out an available system which looks appropriate in the classroom. He might find that, with modification, it will serve his need. It is more likely that he will find he must develop his own system. (p.504)

Although the description of the prototypes for the initial selection of teachers may betray a number of assumptions about the interaction of such treatments and the student characteristics described in the previous chapter, it was considered unwise not to consider other aspects of teacher behaviour that might affect the outcome. Accordingly a number of observation systems were examined to see if one might meet the needs of the present investigation and to see what other researchers were measuring.

In most category systems units of behaviour (brief time units, say five-second intervals, or natural units such as the single statement, question or direction) are classified into a limited number of categories. It is important that the categories are mutually exclusive, inclusive of all the observed behaviour, conceptually fruitful, and can be used by observers with high reliability; and that the techniques of recording provide an adequate and representative sample of classroom behaviour.

Not all the behaviour in the classroom can be observed at the one time. The eye and the camera can only receive stimuli from a limited field of view, and although the ear and the microphone are in this sense multi-dimensional they have greater localization, threshold, and interference problems. The analogy of focussing attention is a good one: observation is selective and brings into perceptible relief only a fragment of the total observable (and audible) behaviour. As there is considerable overlap among categories and terminology of the various systems, it is important that the viewpoint of the investigator who developed each instrument be recognized, particularly where there must be a considerable degree of interpretation before an act is recorded. It is known that eyewitnesses give different accounts of the same incident, so that it is expected that observers would need to be trained to adopt the same viewpoint to achieve high reliability in recording classroom behaviour.

Many observation schedules and category systems are described in Raths, Pancella, and van Ness (1967), Simon and Boyer (1967), Hyman (1968), Nelson (1969), and Dunkin and Biddle (1974). A large number of these was examined and a smaller number considered in detail. This smaller number is discussed in the following section, and in section 6.6 the development of a derivative Category system for the Analysis of Teaching Styles (CATS) is described.

Among empirical studies which have analyzed classroom behaviour in terms of social-emotional climate are those of Anderson and his colleagues (1945, 1946a, 1946b), Withall (1949), Flanders (1965), Hughes (1962), Perkins (1964), and Amidon and Hunter (1967).

Anderson and his colleagues (1945, 1946a, 1946b) produced a demonstrably reliable observation schedule for the categorizing of teacher and pupil behaviour, and investigated the effects of 'dominative' and 'integrative' teacher behaviour on pupil behaviour. Withall (1949) developed a system for classifying records of teachers' verbal behaviour into seven categories, which ranged along a continuum from pupil-centred statements, through neutral and problem-centred statements, to teacher-centred statements. A comparison of the proportion of statements falling into the various categories can give an index of the social-emotional climate of the classroom, and teachers can be characterized by the pattern of their statements.

As did Withall (1949), Flanders (1965) assumes that the teachers' verbal behaviour constitutes an adequate sample of his total pattern of influence, which also implies that his verbal behaviour will be consistent with his non-verbal behaviour. In his system, at the end of each three-second interval the dominant verbal behaviour of the interval is classified into one of 10 categories: (1) accepts student's feeling, (2) praises or encourages, (3) accepts or uses ideas of student, (4) asks questions, (5) lectures, (6) gives directions, (7) criticizes or justifies authority, (8) student talk-response, (9) student talk-initiation, and a residual category, (10) silence or confusion. The original sequence of behaviours is preserved on the record for later analysis. Flanders found that classroom activity tends to fall into relatively homogeneous periods of activity, such as introducing new material, evaluating homework, testing, and so on. As a consequence he found it advantageous to make separate observation records for these periods. His principal hypotheses concerned the effects of 'direct' and 'indirect' teacher influence. Categories 1, 2, 3 and 4 were regarded as containing behaviours showing indirect influence, and categories 5, 6 and 7 as containing behaviours showing direct influence. He found that a sustained above-average level of direct influence restricts learning and produces less desirable attitudes, and that students who achieved most and had the more favourable attitudes were in classes exposed to more flexible patterns of teacher influence. This finding has been supported in a number of studies (see Soar, 1968). The Flanders Interaction Analysis Categories (FIAC) has been widely used as a research instrument in its original form, in slightly modified form, and in a form in which some of the categories were further subscripted to give greater detail, but in which the data could be 'collapsed' back into the original 10 categories. Flanders (1970) reported 7 completed projects of his own, and 16 by other investigators which have associated some measure of pupil learning or behaviour with an interaction analysis variable based on the FIAC. Evans (Note 1) reported 10 studies in science education between 1966 and 1969 which used FIAC, 3 mentioned previously by Flanders.

Hughes (1962) views the teacher as performing a number of functions for the pupil: (1) controlling functions, (2) imposition of teacher (sic), (3) facilitating functions, (4) functions that develop content, (5) functions that serve as response, (6) functions of positive affectivity, and (7) functions of negative affectivity. The verbal and non-verbal behaviour is classified accordingly. Use of Hughes' Provo Code has indicated that teachers demonstrate different patterns of teaching, and that these different patterns do affect the learning of children in different ways. In contrast with the findings of Medley and Mitzel (1958a), Hughes found that teachers' behaviour patterns were relatively stable over time.

Perkins (1964), building on the category systems of Flanders and others, devised a system which requires one observer to classify teacher behaviour and teacher role, while simultaneously another observer classifies student behaviour and student role. Using recording instruments which moved a recording tape at constant speed, the time scale was controlled, and it was found possible to categorize and record from 5 to 15 observations per minute.

In a sense most recent systems of interaction analysis are heirs to the approaches, procedures and insights of previously developed systems, although some more closely resemble earlier systems than others, both in design and purpose. Such systems are not necessarily any the less significant for being more obviously derivative. The very useful Verbal Interaction Category System of Amidon and Hunter (1967) is derived from the FIAC. The 12 categories are grouped according to whether they involve teacher-talk or pupil-talk and whether this talk is initiatory or responsive. In essence, the system emphasizes the distinctions between 'narrow' and 'broad' questions and pupil interactions with the teacher and with other pupils. Such modification of an established system is encouraged by Flanders who sees the orthodoxy of fixed category systems as "inhibiting creative applications and innovations" (Flanders, 1970, p.124). He then illustrates how, depending on the purpose of the observation, categories can be subscripted to provide greater detail, how multiple-coding procedures can increase the flexibility of an instrument, and how different displays of data can assist in the analysis by emphasizing certain comparisons.

Other studies view classroom interactions in terms of communications. The investigation of Lewis, Newell, and Withall (1961) was concerned with the development of a category system for the analysis of classroom interaction, and a test of its sensitivity to the communication patterns of two teachers using dissimilar teaching methods, a 'case-study' method and a 'student-centred' method. From the verbal behaviour, and the non-verbal when detectable, the observer was asked to infer the intent of each act of communication. One record was made for each 10 seconds, or more frequently if there was a shift in intent. The Bales (1950) Process Analysis categories were revised to provide categories which were more descriptive than evaluative, and to minimize the degree of inference needed by the observer recording classroom interaction. The categories were: (1) asks for information, (2) seeks or accepts direction, (3) asks for opinion or analysis, (4) listens, (5) gives information, (6) gives suggestion, (7) gives direction, (8) gives opinion, (9) gives analysis, (10) shows positive feeling, (11) inhibits communication, (12) shows negative feeling, and (13) no communication during the 10-second interval. It was concluded that the observational categories were sensitive enough to pick up different patterns of communication.

Galloway (1966) developed a system for categorizing non-verbal behaviour. In earlier work he had found evidence that pupils use a teacher's non-verbal behaviour such as facial expressions, movements or actions, and nuances of intonation and inflexion of the voice to check on the 'fidelity' of his verbal statements. He quotes Bernstein (1961) as finding that children from the lower socio-economic classes rely 'almost exclusively' on the non-verbal for the detection of meaning in the classroom. Galloway's categories are grouped under two headings, (1) those designated as encouraging communication (enthusiastic support, help, receptivity, and a neutral category, 'proforma') and (2) those designated as inhibiting communication (inattention, unresponsiveness, dissapproval). His unit of analysis is the communicative act.

Joyce and Harootunian (1967) have developed a coding system related to 'conceptual systems' theory (Harvey, Hunt, & Schroder, 1961) which classifies teacher behaviour in non-evaluative patterns or styles. The coding system contains four principal categories: (1) sanctions, (2) procedures, (3) information, and (4) activity initiation, each of which has in turn a number of subdivisions. For example, within the 'Information' category the subdivisions are (I-1) helps child evaluate information, raise hypotheses, make inferences, define or advance problem, (I-2) helps child find information, (I-3) asks lecture question, (I-4) makes lecture statement, and (I-5) gives conclusion.

Another emphasis has been on the analysis of cognitive processes in the classroom. Smith and Meux (1962), Bellack, Kliebard, Hyman, and Smith (1966), and Taba, Levine, and Elzey (1964) have given much consideration to the difficult problem of the definition and identification of the unit of thought in classroom discourse. Smith and Meux, working from records of class lessons, detected two main types of classroom discourse, the episode which consists of exchanges among speakers, and the monologue. Within each form of discourse they further analyzed statements according to their logical nature and devised the following categories: (1) defining, (2) describing, (3) designating, (4) stating, (5) reporting, (6) substituting, (7) evaluating, (8) opining, (9) classifying, (10) comparing and contrasting, (11) conditional inferring, and (12) explaining, to which was added another category having no logical significance, (13) directing and managing classroom. Smith and Meux suggest that the pattern of these verbal behaviours will vary from teacher to teacher and from one area of content to another.

Drawing on the previous work of Smith, and on Wittgenstein's identification of language games, Bellack et al. (1966) categorized the verbal interplay of teachers and pupils into four 'pedagogical moves': (1) structuring, (2) soliciting, (3) responding, and (4) reacting, and found that these occurred in cyclical patterns which they described as 'teaching cycles' after the manner of Smith

(1960). According to its meaning, verbal communication was further classified as having (1) substantive meanings (the specific content and concepts of the lesson), (2) substantive-logical meanings (those cognitive processes involved in dealing with subject matter: defining, explaining, and so on - these can be seen to be a modification of the logical categories of Smith and Meux), (3) instructional meanings (the social-managerial aspects of the classroom which are a part of teaching: assignments, routine procedures, and so on), or (4) instructional-logical meanings (didactic processes such as giving directions, explaining procedures, rating). This study provided a picture of the remarkable similarities among teachers and supports the position taken by Flanders, Amidon, Hughes, and Perkins that the teacher has a clearly defined role as initiator and dominant authority. Analysis of teaching cycles showed that the basic pattern of verbal interchange was the soliciting-responding pattern, and that teachers tended to maintain a consistent emotional style over time.

Gallagher and Aschner (1963) developed a category system primarily on the operations of the intellect as proposed by Guilford (1956). Working from tape-recordings of lessons, the thought processes of gifted children were inferred from their verbal responses and any elucidating details that could be added by observers of the lesson. These thought processes were then classified into the categories (1) routine, (2) convergent memory, (3) convergent thinking, (4) evaluative thinking, and (5) divergent thinking. They found that the profile of the divergent production for both boys and girls followed the same general pattern as that of the teacher, but noted that a relatively small proportional increase in the teacher's divergent questions brings forth a large increase in the divergent production of the students. Gallagher (1967) further developed this category system to indicate (1) the level of conceptualization (data, concept, generalization), (2) the style of thinking (description, expansion, explanation, evaluation), and (3) the emphasis of the teacher on skills or content development. Using the instrument on six biology teachers teaching the same topics with the same BSCS Blue Version materials to high ability students, he found substantial differences among the teachers on all three dimensions of behaviour: "each teacher filters the material through his own perceptions" (p.17).

Davis and Tinsley (1967) developed a Teacher-Pupil Question Inventory based on the categories of Bloom's (1956) Taxonomy of Educational Objectives: Handbook 1, Cognitive Domain, and on Sanders' (1966) adaptation of it. Questions asked during the lesson are judged on their form and inferred intent as well as on the nature of the response they elicit and their reception by the pupil or teacher. Davis and Tinsley found that acquisition of knowledge was the dominant objective apparent in the questions of both teacher and pupil.

Among investigations of classroom interaction in secondary school biology are those of Parakh (1968, 1969) and the joint work of Evans (1969) and Balzer (1969). Evans and Balzer collaborated to develop their observation system and to collect data; however, they carried out separate analyses and research studies based on this data. Both research teams found shortcomings in the available techniques of classroom analysis of biology, which may have long periods of significant activity such as laboratory work characterized by comparatively little verbal interaction. In their investigations they found evidence of considerable non-verbal behaviour that was 'pedagogically relevant'. Both teams have developed a number of subcategories which permit more detailed analysis than is usual of the content development and cognitive aspects of the classroom. Balzer and Evans have focussed entirely on the recording of teacher behaviour to develop their Biology Teacher Behaviour Inventory, and Parakh, although placing major emphasis on the behaviour of the teacher, also included in his schedule categories for the verbal behaviour of pupils which permit analysis of teacher-pupil interaction patterns.

Parakh developed his categories from an analysis of tape recordings of classroom lessons, replayed as often as necessary, and of notes made at the time, particularly of the non-verbal behaviour of the teacher. His major categories were (1) evaluative dimension (affective-substantive), (2) cognitive dimension (giving or seeking substantive information), (3) procedural dimension, (4) pupil talk dimension (giving or seeking information), and the residual categories,(5) silence, and (6) not categorizable in the above system. The information-giving-and-seeking categories were sub-divided into a number of logical and quasi-logical operations such as stating facts, defining terms, explaining, and evaluating. Among Parakh's findings were: that teachers talked about 75% of the time in the classroom and about 50% of the time in the laboratory, and that the giving and seeking of information accounted for 54% of the classroom time and 42% of laboratory time. His data suggest a preponderance of the lower levels of cognitive activity as described by Bloom (1956).

Evans (1969) and Balzer (1969) developed their instrument in a similarly inductive manner, but used videotaped records of class lessons to identify and record an accurate and comprehensive account of classroom behaviour. The categories of their Biology Teacher

Behaviour Inventory (BTBI) and a glossary of teacher behaviour were developed from this list. Behaviour was classified into one of four expressive forms, verbal, non-verbal, congruent, and contradictory. Congruent behaviour was defined as "simultaneous verbal and non-verbal behaviour in agreement, i.e., one set of behaviors was used to stress, reinforce or amplify the other" whereas contradictory behaviour was defined as "simultaneous verbal and non-verbal behaviors inconsistent with one another" (p.222). These two categories are similar to Galloway's (1966) grouping of non-verbal behaviour into that encouraging communication and that inhibiting communication. Interestingly enough, Evans' study shows that the proportion of the total recorded behaviour of his eight teachers that could be regarded as contradictory was only 0.1%, which gives some support to Flanders' (1965) assumption that a teacher's non-verbal behaviour is consistent with his verbal behaviour. Flanders' other assumption, that a teacher's verbal behaviour can be taken as an adequate sample of his total pattern of influence, however, may be challenged by Evans' finding that there is a slightly greater proportion (38.9%) of purely non-verbal expression in biology classrooms than purely verbal expression (34.8%). Total verbal behaviour, that is expression categorized as either verbal or congruent, accounted for approximately 61% of the classroom time (Parakh's figures were 75% for the classroom and 50% for the laboratory). Teacher behaviour was further classified according to its inferred intent into the following major categories: (1) management, (2) control, (3) release, (4) goal setting, (5) content development, (6) affectivity (positive and negative) and (7) undecided. Content development was either teacher-centred (5a) or pupil-centred (5b), and further sub-divided into (5-1) procedures, (5-2) knowledge, (5-3) scientific process, (5-4) tentativeness of knowledge, (5-5) generalizations, (5-6) articulation of content, and (5-7) facilitation of communication. Each of these sub-divisions could be communicated in one or more of the following ways: (a) states, (b) asks, (c) shows, (d) acknowledges, and (e) clarifies. Despite the apparent complexity of the categorization, Scott's (1955) coefficients of 0.92 and above were obtained for inter-observer agreement. Evans found that nearly all teacher behaviour was concerned with management and teacher-centred content development (Balzer confirmed that this was as true for teachers of BSCS courses as for teachers of other biology courses). As Parakh (1968, 1969) found also, most of the content development was concerned with the giving and receiving of facts, definitions, and terminology.

In both Parakh's study and the work of Evans and Balzer teachers assumed their natural teaching styles and in each study wide variations in teaching behaviour were found; so although the samples were small (eight teachers for Parakh and 11 teachers for Evans and Balzer), the close correspondence of their findings suggests that the picture they paint of biology teaching in the United States of America may be a good likeness.

Although all of the above systems had their attractive features none of them appeared to meet the needs of the present investigation. And although the development of yet another category system was not the main purpose of the present research, it was felt necessary to develop one which would incorporate those elements of teaching style which might be expected to interact with student attributes in the context of this particular investigation. In addition, the system needed to be usable with a minimum of prior training. Because of this, the system necessarily used fairly obvious distinctions which required little interpretation. Many of the pervasive subtleties of teaching style are unlikely to be measured by such a system but some may be. It was hoped that the category system would provide a reflection of teaching style which was sufficiently discriminating between teachers and reliable in its operation to verify that there were two different instructional treatments, and to allow the treatments to be described in useful detail.

6.6 THE DEVELOPMENT OF A CATEGORY SYSTEM

Balzer (1969) and Evans (1969) argue strongly in favour of a category system inductively derived. They propose that "every teacher verbal and nonverbal classroom behavior perceived as influencing the teaching-learning situation" (Evans, p.221) be recorded, that no arbitrary exclusions be made, and that the resulting comprehensive list of behaviours be then sorted into a limited number of categories. The deductive derivation of category systems based on theoretical models of teaching, they suggest, can lead to difficulties in coding behaviours which had not been considered in the preconceived model. Allon (1969) sees a danger that the categorical structure can impose the system's deductive definition of behaviour on the behaviour to be observed. Although Evans' point is well taken, the so-called inductive method does not meet Allon's objection any better than the deductive method. The development of any category system depends on the investigator's perception of similarities among the objects or events. As Hanson (1961) points out, it is a theory-laden undertaking. A system of mutually exclusive and logically related categories requires a theoretical perspective to provide criteria of relatedness. In a deductively derived system the assumptions and orientations of the investigator are usually made explicit, but there is a danger that they may be assumed to play no part in the inductively derived system. Evans and Balzer state that only behaviour "perceived as influencing the teaching-learning situation" will be classified, and then "according to the inferred intent of the behaviour" (Evans, p.222). The freedom from positional judgements may be more apparent than real. May and De Vault (1967) used a factor analytic procedure which is therefore partially free from subjective grouping of variables, but then had some difficulty in assigning useful descriptive labels to the factors they isolated. In this investigation a category system is deductively derived. Whether or not this leads to later difficulty in coding behaviours not anticipated will be reflected in the number and complexity of the ground rules needed to deal with 'difficult' behaviours, and the proportion of behaviours classified as non-categorizable. It is impossible to be entirely free from idiosyncratic perceptions of reality: the investigator can only hope to be open-minded when it comes to looking at the data, and to let the data speak for themselves as much as they can.

The first step in developing a classroom observation instrument for this investigation was to decide on what features of classroom behaviour to focus attention. A classroom observation system attempts to produce an accurate and objective record of some aspects of classroom behaviour by largely ignoring others. Classroom behaviour is far too complex for all incidents and their immediate context to be distinguished and recorded. Attention is focussed according to the investigator's assumptions of the relevance of some dimension of classroom behaviour rather than another to the purpose of his study. For example, if all solicitation of response by the teacher is to be coded under one symbol then the distinction between questioning and direction is lost once the act has been recorded. What is gained is an accurate frequency of behaviour solicited by the teacher, as distinct from pupil-initiated behaviour. It is important that the system does not lose information more important than that it retains. The development of a category system designed to preserve certain distinctions in classroom behaviour is greatly facilitated by the opportunity of repeated examination of videotape or audiotape records of classroom interaction.

The development of the skills of inquiry and of critical thought are basic objectives of the Web of Life biology course, and so it was decided to concentrate on those distinctions between teaching styles that appeared likely to achieve these objectives and those which were not. As a result of consideration of the observation schedules described in section 6.5 the following distinctions were incorporated into the category system used in this investigation.

Open and Closed Behaviour

Morrison and McIntyre (1969) note that

> the numerous attempts to categorize teachers in terms of the general motivational styles they adopt ... often bear a close relation to the distinction between practices which emphasize learning through discovery and self-directed activity, and learning in closely directed and extrinsically motivated conditions. (p.133)

Among the many distinctions utilized by previous research are 'authoritarian/democratic', 'dominative/integrative', 'teacher-centred/pupil-centred', 'direct/indirect', 'convergent/divergent', 'open/closed', 'directive/non-directive'. The distinctions between some of these pairs are quite clear, but between other pairs the distinction rests finally with the way in which individual acts of classroom behaviour are coded. The studies of Gallagher and Aschner (1963) and Hughes (1962) indicate that an increase in 'divergent' or 'open' behaviour by the teacher has corresponding effects on pupil behaviour. Stern's (1963) comprehensive review of studies of 'non-directive' and 'directive' instruction indicates that the majority of researchers have found no particular advantage for either approach in terms of cognitive achievement, but points to the work of Wispé (1951) and Patton (cited in Stern, 1963) which suggests that pupils work best in situations that conform to their preference; for example, pupils who reject traditional sources of authority tend to prefer student-centred instruction. The research on creative thinking, reviewed by Getzels and Madaus (1969), the studies on 'dogmatism' by Ehrlich (1961), Fillenbaum and Jackson (1961), and Mouw (1969), and the evidence of classroom interaction studies on social-emotional climate, all provide support for considering the distinction between open and closed behaviour as potentially fruitful in the analysis of classroom interaction.

Rokeach (1954, 1960) has provided the general theoretical framework for the consideration of open and closed belief systems. The closed-minded or dogmatic individual characteristically believes that he knows the answers, or that they are known by the appropriate authority,and that his outlooks are the appropriate ones to hold. The open-minded individual is characteristically less certain of his omniscience and is prepared to tolerate challenging evidence and other points of view.

Open-minded and closed-minded individuals may react very differently in the teaching and learning of concepts and skills, particularly where the development of inquiry skills is seen as a challenge to authority. The terms 'open' and 'closed' seem to be appropriate labels for teacher and pupil behaviour and are, for the purposes of the present investigation, characterized as follows. Open behaviour on the part of the teacher accepts the pupil's frame of reference; and encourages pupils to question, to develop critical capacities and skills in inquiry processes, and to be creative. It is designed to lead a pupil to find out for himself rather than to accept another's word, and is expected to promote pupil-initiative and independence of thought and action. In contrast, closed behaviour exhibits the 'conservative' view, is authority-centred, and is designed to promote uncritical acceptance of authoritative accounts of the present state of knowledge and of methods and procedures. Pupils are encouraged to adopt the dominant values and outlooks of adult society as interpreted by the teacher. Although some educational theorists (Weir, 1963) have strongly advocated the open position, the descriptions of open and closed behaviour given here are intended to be non-evaluative.

Positive and Negative Reinforcement

There is, however, another class of activities which are not appropriately categorized as open or closed, but which have a discernible affective connotation and contribute to the social-emotional climate of the classroom by eliciting or reinforcing various kinds of pupil behaviour. The teacher may smile, laugh, joke, praise, encourage, or otherwise show approval: such behaviour should be categorized as showing positive affect. On the other hand, the teacher may frown, reprimand, accuse, admonish, threaten, use sarcasm, or otherwise show disapproval: such behaviour should be categorized as showing negative affect. Evidence from the work of Flanders (1965) and Perkins (1964), working with teachers of grade 7 and grade 5 pupils respectively, indicated that the proportion of teacher behaviour in the classroom which could be regarded as praise or encouragement was of the order of 2%, and the proportion of behaviour that could be regarded as correction, criticism or rejection was between 0.5% (Perkins) and 4% (Flanders). Parakh (1968, 1969), Evans (1969), and Balzer (1969), working with teachers of high school biology (10th grade), put the proportions at 0.5% to 1.4% for positive affectivity and 0.8% to 1.3% for negative affectivity. Thus the proportions of teacher behaviour which fall into these categories are small. However, the literature reviewed by Glaser (1969) and Rosenshine (1971) attests to the importance of positive and negative reinforcement in learning, and the importance of such behaviour may be far greater than the frequencies tend to suggest. This may also be a dimension on which patterns of behaviour are relatively stable for individual teachers and yet which vary considerably from teacher to teacher; that is, this dimension may be important in categorizing teaching styles.

Managerial and Substantive Functions

An analysis of the data provided in the studies of Hughes (1962), Perkins (1964), Parakh (1969), and Balzer (1969) indicates that high frequencies of teacher behaviour occur in both 'classroom management' and 'content development' categories. Although many category systems do not make this distinction, it is possible that while a teacher may use one pattern of behaviour for dealing with substantive content, he may use another pattern for dealing with class organization. The use of different patterns of behaviour for different purposes has been called flexibility, and a number of workers using FIAC (e.g., Moskowitz, 1967; Flanders, 1970;

Trebilco, 1975), have found it to be an important characteristic of teachers' individual teaching styles.

All pedagogically relevant behaviour can initially be considered as 'managerial' or 'substantive'. 'Substantive' behaviours are primarily concerned with the presentation and development of subject matter in the classroom and the acquisition of related skills and attitudes. These behaviours are based on efforts to achieve the basic teaching objectives, which may be as much concerned with the attitudes and the procedures of inquiry (skills) as with the products of inquiry (content). Teachers vary considerably in their integration of these objectives. Some teachers appear to spend most of their time concerned with content and skills appear to be developed incidentally, whereas other teachers appear to emphasize skills development, the content serving mainly as the vehicle. In practice it is often very difficult to separate the two (Bingman, 1969). 'Managerial' behaviours are those which are primarily concerned with the organization of the instructional environment, such as routine organization of the classroom and laboratory, control of student behaviour, and the establishment and maintenance of a congenial classroom climate. Like Gallagher and Ashner's 'routine' category it includes all acts of positive and negative reinforcement.

Initiation and Response

One important decision was whether to focus attention primarily on (a) the logical nature of the statements; that is, on categories similar to the 'logical categories' of Smith and Meux (1962) (defining, describing, designating, stating, reporting, substituting, evaluating, opining and so on), (b) the level of cognitive operation being used, such as the 'level of conceptualization' (data, concept, generalization) used by Gallagher (1967) or the modification of Bloom's (1956) categories used by Davis and Tinsley (1967), or (c) the functions that the statements (or acts) serve, for example, the 'pedagogical moves' of structuring, soliciting, responding, and reacting delineated by Bellack et al. (1966). Flanders (1970, p.176) asserts that logical functions in classroom discourse are very difficult to code if all logical processes are to be taken into account. It is significant that the studies which emphasize the logical nature of classroom interaction, such as those of Smith and Meux (1962), Bellack et al. (1966), and Gallagher and Aschner (1963) rely on unlimited access to transcripts, audio or videotape records for their analyses. Even then it appears that there is a rich variety of logical patterns depending on the immediate context, so this approach may not be the most appropriate for the categorization of teaching styles. A multiple-coding technique with unlimited access to videotape records can, of course, code a single statement or act on each of these dimensions, but of the three, the pedagogical moves were judged to be the most appropriate orientation for this study in that they are more directly suitable for the analysis of the balance of teacher initiative and response. This balance, together with the balance of open and closed behaviour, provides a basis for the quantitative description of what in this investigation is called teaching style, that is those elements of teacher behaviour that appear less likely to be determined by the conscious adoption of a particular teaching strategy.

For the present investigation, the broad categories of 'initiation' and 'response' will suffice for the categorization of pupil behaviour, but a more detailed analysis of teacher activity is required. Initiatory moves have been sub-divided into 'presentation', 'questioning', and 'direction' and responsive moves have been subdivided into 'development' and 'evaluation'. These five categories of teacher behaviour are now defined.

Presentation includes all instances where new information is given in a way that is not immediately concerned with the various possibilities of response. Like Bellack's category of structuring it commonly launches a new teaching cycle (Smith, 1960; Bellack et al., 1966). The information can be fact or opinion, and may deal with knowledge, attitudes, or skills. However, if information is given as a result of a request, this is to be considered as development.

Questioning is another initiatory move and includes behaviours which are intended to elicit an answer. Some questions ask for information about content and procedures, and require the use of memory alone; others call for the use of higher-order intellectual skills: Comprehension, Application, Analysis, and cognitive Evaluation (Bloom, 1956). Other questions do not presume that there is a known or right answer and are regarded as open or divergent (Guilford, 1956) questions (cf. Wright & Nuthall, 1970); they may require the expression of an attitude, the proposal of a plan, or some other creative response. Most solicitations requiring verbal or written response are included in this category; rhetorical questions are not. Some imperatives such as "Tell me how many legs an insect has," are regarded as de facto questions. So, questions are broadly categorized as being memory questions, higher-order questions, or open questions. The memory questions include the Gallagher and Aschner (1963) categories of routine, and convergent memory questions, and similarly the higher-order questions include the Gallagher and Aschner categories of convergent thinking and evaluative thinking questions.

Direction includes behaviours which are intended to structure the behavioural response, either in the expectation of compliance as when a command is given, or by providing alternative courses of action as when a suggestion is made. Most solicitations requiring a non-verbal response are directions. Requests, which typically begin with the formal interrogative "Would you please ...?" are regarded as de facto directions. Any active invitation for another person to say or do something is to be regarded as either a question or a direction. Some directions give clear positive or negative reinforcement.

Development is a response behaviour and includes those acts which serve to modify what has gone before, by clarifying, interpreting, summarizing, expanding, or synthesizing. Development takes place in the context of what has gone before, but need not be directly elicited by the preceding event. Unlike Bellack's 'reacting' category, it does not generally include evaluative responses. The evaluation of, or an opinion about, substantive content or processes is included, being similar in function to the other developmental operations, but the evaluation of pupil behaviour is considered in a separate category.

Evaluation of pupil behaviour includes behaviours that could be classified as Evaluation in Bloom's (1956) Taxonomy of Educational Objectives, and also most of those behaviours that are categorized as positive and negative affect in other systems (Lewis, Newell, & Withall, 1961; Flanders, 1965; Hughes, 1962; Amidon & Hunter, 1967; Evans, 1969; Parakh, 1969; Hough & Duncan, 1970). Judgements of the correctness or appropriateness of a pupil's response are sometimes made without any discernible emotive connotation, usually in terms of external or public criteria (such as accepted definitions, conventions, or empirical verification). These are what Hough and Duncan (1970) would describe as confirmation and corrective feedback and are examples of closed (cognitive) evaluation. Verdicts usually indicate closed behaviour; open behaviour is much more tentative. Where the judgements have discernible overtones of personal criteria, and could be described as 'encouragement, praise, or approval' or 'criticism, reproof, or disapproval' categorization is in terms of positive reinforcement and negative reinforcement.

The teacher's non-verbal behaviour is considered separately. Flanders (1965) regards a teacher's verbal behaviour as a representative and adequate sample of his total behaviour. However, Galloway (1966), Parakh (1969), Evans (1969), and Balzer (1969) all demonstrate the importance of taking both verbal and non-verbal behaviour into account. Where they accompany verbal behaviour, gestures, body movements, vocal nuances, and facial expressions all help to elucidate the intent of the verbal behaviour. Evans (1969) provides evidence that almost all non-verbal behaviour in his sample was congruent with the teachers' verbal behaviour so there is some support for Flanders' view that a teacher's verbal behaviour is a representative sample: it may not however be adequate. In biology lessons there may be large periods of time with no verbal interaction between teacher and students when they are working on laboratory investigations. The teacher may write on the chalkboard, use other audio-visual media, assist with the focussing of a microscope, or walk around the class ready to assist with enquiries: all instructionally-significant behaviours. In this investigation the observer made a written account of all such behaviour during periods of silence, and noted whether it was open, closed or audio-visual in nature. He also noted any instances of contradiction between verbal and non-verbal behaviour such as sarcasm. Contradictory non-verbal behaviour modifies the categorization of the accompanying verbal behaviour, usually an evaluation or a direction.

> In speech we have available a subtle and extensive resource of stress, and change of pitch, of voluntary manipulation of pause and speed and volume, of change in voice quality, to say nothing of the nod and the wink that can modify, or even flatly contradict, the meaning of what we say.
> (Brazil, 1969, p.83)

To ensure that the category system is inclusive of all possible behaviours, there must be a residual category for behaviour that is pedagogically irrelevant, such as 'teacher talking to visitor'; for when there is confusion or many pupils talking at once; or when for some other reason the behaviour cannot be categorized. In periods of silence the teacher's non-verbal behaviour usually serves some instructional function, only instructionally non-functional silence should be regarded as non-categorizable. With the availability of annotated transcripts and original audiotape records, it is hoped that most events initially recorded in this category can on reconsideration be placed elsewhere. If the categories are clearly defined and mutually exclusive, if the categories are inclusive of all instructionally functional behaviour, and if the teachers and classes are going about their work in a sincere and business-like fashion, the proportion of entries in this category should be small.

Teachers' behaviour can also be differentiated on the basis of their preferred group size for teaching. Some teachers spend most of their time interacting with students on an individual or small group (2-4) basis, whereas others spend more time interacting with the class as a whole. The Web of Life course emphasizes individual and small group work and the degree to which teacher-pupil interaction takes place at this level may reflect basic aspects of an individual's style rather than conscious strategy.

Using the distinctions described in this section, a Category System for the Analysis of Teaching Styles (CATS) was developed. At the first level of categorization a unit of behaviour can be classified into one of nine basic categories, viz., teacher behaviour (non-verbal, development, evaluation, presentation, questioning, direction), pupil behaviour (response, initiation), and non-categorizable. Having been so located the unit of behaviour can then be categorized at a second level. Non-verbal behaviour (category 1) can be further categorized as open (O), closed (1), showing positive affect (+), negative affect (-), or being audio-visual (a) in nature. Development (category 2) and presentation (category 4) can be either open (O) or closed (1). It is expected that open behaviour will be rather rare. The more usual behaviour mode would be closed, although this term is not meant to have pejorative overtones and is recorded by the neutral tally mark (1). Evaluation (category 3) can be recorded as having positive affect (+), negative affect (-), or no discernible affective overtones (1). Questions (category 5) can be memory questions (1), higher-order questions (h), or open questions (O). Directions (category 6) can show positive affect (+), negative affect (-), or be open (O), or closed (1), in nature. Pupil behaviours (categories 7 and 8) similarly can be further categorized as (+), (-), (O), or (1). As a third level decision, entries in categories 1 to 6 can be further subscripted with an 'i' if the teacher is interacting with an individual or small group of students. At a fourth level, the same unit of behaviour can then be categorized as being concerned with managerial functions (M) or substantive content and skills (C), except that all behaviour classified as having positive or negative affect is regarded as managerial in function.

Non-categorizable behaviour is simply entered in category X as a tally mark (1).

Avoiding the logically-incompatible and conceptually-improbable permutations in the system, a single behaviour can be classified in one of 77 different ways. Flanders (1970) observes that any practical classification system suffers a loss of categories not only because of logically-incompatible permutations, but also because the range of natural behaviours that occur in the classroom is limited due to the rather crude and inclusive nature of most categories. Nevertheless, a ninefold increase in the number of categories as a result of adding additional clusters and subscripts suggests that a decided qualitative improvement in the description of classroom behaviour might also follow. Where the permutations of a multiple coding scheme are restricted as described above, the system would be described by Flanders (1970) as a semi-restricted multiple-coding category system. The categories of CATS are shown in Table 6.6.1 arranged in decision levels.

Unit of Analysis

Before trial with direct classroom observation, one further tentative decision was made: about the unit and frequency of behaviour to be coded. Medley and Mitzel (1963, p.300) prefer the natural unit: each occurrence, each statement, each action is recorded. They contend that interpretation is easier when it is possible to refer to the number of acts, or proportion of statements. However, if a teacher asked an open question for five seconds and made a closed statement lasting 20 seconds, the situation could not be regarded as accurately described by saying that one open question and one closed statement were noted.

TABLE 6.6.1

DECISION LEVELS OF THE CATEGORY SYSTEM CATS

First Level Categories			Second Level Categories						Subscript	Fourth Level Categories	
			Open	Closed	Positive	Negative	Higher-order	Audio-visual	Individual	Managerial	Substantive
1	Teacher	non-verbal	0	1	+	-		a	i	M	C
2		development	0	1					i	M	C
3		evaluation		1	+[b]	-[b]			i	M[b]	C
4		presentation	0	1					i	M	C
5		questioning	0	1			h		i	M	C
6		direction	0	1	+	-			i	M	C
7	Pupil	response	0	1	+	-				M	C
8		initiation	0	1	+	-				M	C
X	Non-categorizable										

[b] Positive and negative reinforcement can only be categorized as M at the fourth level decision.

Those who advocate the use of a time-unit suggest that a time-unit be chosen that closely corresponds to the average time taken by a natural unit of behaviour in the classroom: short enough so that separate behaviours are distinguished from one another and long enough so that the same behaviour is not regularly coded in successive intervals. On this point experience appears to vary, and the time-units used range from 15 seconds (Wright, 1959) down to three seconds (Flanders, 1965). Wright, recognizing the limitations of any fixed division in approximating the observation of natural behaviours, determined empirically that the 15-second division at alternate intervals provided a valid sample. Evans (1969) decided to use a 10-second interval finding that many teacher behaviours were not completed in five seconds and that a large number of behaviours were lost if a 15-second interval was used. Parakh (1969) observing biology lessons, as did Evans, established by trial a time-unit of five seconds for his system.

In this study a compromise has been adopted. Written transcripts were made from the audiotaped records of classroom lessons and this transcript was divided into five-second segments. Each five-second segment was categorized but if more than one distinctly different behaviour were shown in a five-second interval, each such unit of natural behaviour was categorized. In this way the ebb and flow of rapid pupil-teacher interchange is preserved and yet statements which last over five seconds receive additional weighting through being categorized two or more times.

6.7 RECORDING AND CODING PROCEDURES

Three main methods of obtaining a satisfactory record of class lessons were tried: audiotape record with fixed microphones, videotape record with fixed microphones, and audiotape record with a radio-microphone. Because biology teachers often move into the preparation room, or in the school garden, fixed microphones and videotape were found to be impractical, although a videotape record with fixed microphones and multiple cameras was found to give a splendid record of lessons restricted to the classroom. It was decided to use an audiotape record with the teacher wearing a radio-microphone, but additional measures had to be taken to produce a satisfactory record of the teacher's non-verbal behaviour. The observer had to make a simultaneous record of all non-verbal behaviour that would modify the intent of the teacher's spoken word (e.g., sarcasm, approval, disapproval) and describe the teacher's behaviour during periods of silence.

Annotated audiotape records were made of five representative classroom lessons of 40-60 minutes duration for each of the eight classes in the study. The administration of pretests during the first weeks of the year enabled the investigator to become familiar with the classes and accepted to the point where his presence in the class with radio receiver, cassette recorder, and headphones caused minimal discernible disturbance. In the whole 2000 minutes of recorded classroom activity there were only two allusions to either the presence of the investigator or to the recording of the lessons, and both were accepting and good-humoured. Since the teachers often wore laboratory coats the investigator had a coat specially adapted with the transmitter in an inside pocket and the aerial in a vertical seam. All that was then necessary was for the teacher to don the coat, switch the transmitter on, and place the microphone loop around his/her neck. In this way the teacher's voice was always recorded distinctly, as well as the voices of students engaging in conversation with him. In practice, a small proportion of questions or responses made from the body of the class to the teacher were not entirely audible. This was not considered to be a serious drawback in that the basic purpose was to describe teacher behaviour, and in those cases where pupil talk was not entirely clear, sufficient of the tone or content was discernible to make clear the meaning of the teacher behaviour.

The observer sat where he could clearly hear and see the classroom interaction, particularly the teacher's behaviour, and yet was as unobtrusive as possible: generally at the rear of the classroom. In early trials of the system the observer attempted to keep a timed written commentary indicating such changes in class activity as the use of audio-visual media, attempting particularly to elucidate the pedagogically-significant behaviour, if any, of the teacher during periods of silence. However, when transcripts were made and these initial commentaries were used to categorize the classroom record, it was found that many aural cues picked up by the observer, but unrecorded on tape, were not recorded in the commentary. In subsequent trials the observer wore head-phones and monitored the audiotape record. As a result of the observer being able to hear only what was recorded on tape, the written commentary became a much more complete and satisfactory source of elucidation.

Complete transcripts of each recorded lesson were made with elucidatory comments from the accompanying written commentary included in the transcript in brackets. This transcript was divided into five-second segments by playing the original tape together with a tape which gave an audible signal every five seconds (and a signal of a different pitch every minute) and marking the transcripts in the text wherever this signal occurred. The following extract is taken from a transcript (Lesson 25, p.15) at this stage:

/T: Well take it out Stephen (Insistently)
S: It's just water, I think. Yea, it's just water./
T: Okay. Well we might have lost a bit./
S: (Inaudible remark)
T: Yea, Okay. Now dry it and label it. Alright?
Put it in the incubator./ Now we are right now.
I think we ought to do by now./
S: I don't know. I think my finger's stuck there for
life actually./

Using both the audiotape recording and the transcript, the lesson was then coded on prepared coding forms in the following way. The teacher's first statement is a direction, closed behaviour, directed to an individual, and concerned with the substantive content or skills of the lesson and hence coded in category 6 as 1_i and in column 10 as C. The student's two statements are judged not as responses but as initiatory behaviours, making new observations, and two tallies are entered in column 8 and two Cs in column 10. These three behaviours were entered in one five-second section of the coding form. This 25-second segment when coded would be as shown in Figure 6.7.1.

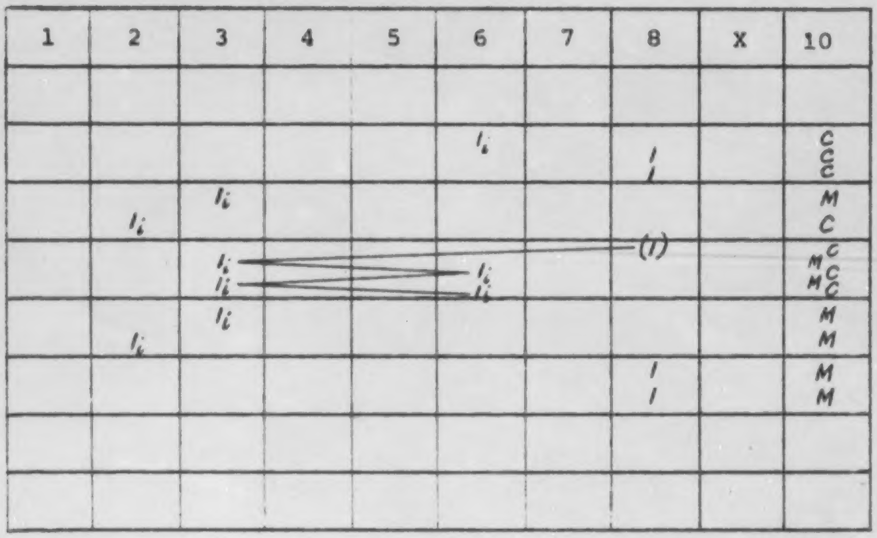


Figure 6.7.1. Portion of encoded lesson

In the third five-second interval it can be seen that there is a rapid verbal exchange. There was insufficient room on the coding form to spread these entries out and the zig-zag line was used to make clear the sequence of events.

Some of the early lessons were fully transcribed, timed, and coded. This took nearly 10 hours for each hour of class time recorded: about five hours for transcription, one hour for timing and up to four hours for encoding. There were a possible 40 hours of lesson time to be processed: a prospect of nearly 400 hours on this task alone and 160 hours of paid labour for an additional coder. This demand was beyond the resources of the investigator and forced consideration of a further sampling of classroom behaviour. To take advantage of the observer's record of non-verbal behaviour, and to enable other coders to provide some checks on reliability, there could be no reduction from the requirement to produce written transcripts sub-divided into five-second intervals. The investigator experimented with a device for producing a signal at five-second intervals which could be recorded simultaneously with the original recordings, but a completely successful method could not be devised in time. Such a device would save one hour per lesson and would provide an accurate time datum on the original tape. So time savings had to be made in the encoding phase. The eight teachers themselves constitute a sample of biology teachers with contrasting teaching styles, and the five recorded lessons are a sample of each teacher's year-long behaviour with that class: what was proposed now was a third-level sampling from the recorded lessons for each teacher. Transcripts of two lessons recorded during the development stages of the CATS system were coded in full, and the results of more economical coding procedures were compared with these. After comparing a one-twelfth, one-sixth, and one-third sampling, it was concluded that a one-sixth sampling provided a substantial saving in coding time and yet did not result in any serious loss from the less common categories of teacher and student behaviour. As a result, every sixth five-second interval throughout the lesson was coded. Meaningful portions of statements lying within the five-second interval which overlapped from the previous interval or into the following interval were coded. Two coders were used in this investigation and the mean frequencies for each category were recorded. In all, the behaviour within 4320 five-second intervals was coded, or 7434 separate units of behaviour. A complete transcript of one lesson is provided for illustration in Appendix III, together with the one-sixth sample coding.

The codings for each lesson were summarized and these summaries were totalled for each teacher. In order to facilitate comparisons between teachers the total lesson time recorded for each teacher was converted to a base of 250 minutes. The frequencies of recorded behaviour in each category were altered in proportion. These data are presented in Tables 6.7.2 to 6.7.10.

TABLE 6.7.2

TOTAL CODED UNITS OF BEHAVIOUR FOR EACH TEACHER

Teacher	Recorded lesson time (minutes)	Coded units of behaviour (one-sixth sample)	Coded units of behaviour (each teacher's lesson time converted to 250 mins.)[a]
1	299	1025	859
2	299	1221	1024
3	262	652	623
4	271	1013	936
5	251	861	861
6	284	912	804
7	231	928	1005
8	263	822	783

[a] The frequencies of behaviour in each category were altered proportionally and, as a result of rounding, each total may not be an exact proportional increase or decrease.

Earlier discussion indicated that a unit of classroom behaviour could be categorized in one of 77 different ways. However, it can be seen that with these classes the samples of coded units of behaviour fell into only 45 of these categories. Six of these 45 are categories of pupil behaviour, but if these are regarded as a result of the classroom environment generated by the teacher and as much a characteristic of the teacher as his direct behaviour (Baird, 1973), then we have eight teachers characterized by scores on 45 dimensions of classroom behaviour.

TABLE 6.7.3

SUMMARY OF CODED BEHAVIOUR : TEACHER 1

		Teacher-individual							Teacher-class								Student			
		1	2	3	4	5	6	Tot	1	2	3	4	5	6	Tot	Tot	7	8	Tot	X
+	M			4				4			1				1	5	7		7	
-	M																			
1	M	1	58	44	1	35	6	145	92	69	5	10	11	33	220	365	64	49	113	
	C	24	67		1	21	6	119		60	3	43	19		125	244	21	28	49	
O	M																			
	C																			
h	M																			
	C				2			2					7		7	9				
a	M								25						25	25				
	C																			
Totals		25	125	48	2	58	12	270	117	129	9	53	37	33	378	648	92	77	169	42

TABLE 6.7.4

SUMMARY OF CODED BEHAVIOUR : TEACHER 2

		Teacher-individual							Teacher-class								Student			
		1	2	3	4	5	6	Tot	1	2	3	4	5	6	Tot	Tot	7	8	Tot	X
+	M			15				15			1				1	16	2		2	
-	M			3				3								3				
1	M	2	79	70	10	77	42	280	26	29	1	32	13	27	128	408	71	28	99	
	C	59	105	2	29	86	5	286		28	1	35	15	1	80	366	66	33	99	
0	M																			
	C																			
h	M																			
	C												15		15	15				
a	M								3						3	3				
	C																			
Totals		61	184	90	39	163	47	584	29	57	3	67	43	28	227	811	139	61	200	13

TABLE 6.7.5

SUMMARY OF CODED BEHAVIOUR : TEACHER 3

		Teacher-individual							Teacher-class								Student			
		1	2	3	4	5	6	Tot	1	2	3	4	5	6	Tot	Tot	7	8	Tot	X
+	M			5			5									5	1		1	
-	M			2			2									2				
1	M	7	15	28	3	29	28	110	138	27	2	18	14	46	245	355	21	21	42	
	C	14	5	1	2	15		37	15	21		57	19	7	119	156	16	6	22	
0	M					1		1								1				
	C																			
h	M																			
	C																			
a	M																			
	C																			
Totals		21	20	36	5	45	28	155	153	48	2	75	33	53	364	519	38	27	65	39

TABLE 6.7.6

SUMMARY OF CODED BEHAVIOUR : TEACHER 4

		Teacher-individual							Teacher-class								Student			
		1	2	3	4	5	6	Tot	1	2	3	4	5	6	Tot	Tot	7	8	Tot	X
+	M			9				9			1				1	10	7		7	
-	M			6				6								6				
1	M		17	60	1	23	33	134	47	45	12	9	37	71	221	355	50	59	109	
	C	14	6		1	9	2	32	1	120	1	75	77	8	282	314	70	49	119	
O	M																			
	C																			
h	M																			
	C					2		2					7		7	9				
a	M							1							1	1				
	C																			
Totals		14	23	75	2	34	35	183	49	165	14	84	121	79	512	695	127	108	235	6

TABLE 6.7.7

SUMMARY OF CODED BEHAVIOUR : TEACHER 5

		Teacher-individual							Teacher-class								Student			
		1	2	3	4	5	6	Tot	1	2	3	4	5	6	Tot	Tot	7	8	Tot	X
+	M			3				3								3	3		3	
-	M																			
1	M	8	48	55	1	33	21	166	66	89	5	47	5	36	248	414	63	55	118	
	C	31	29		1	9	1	71		55		74	15	1	145	216	40	33	73	
0	M																			
	C																			
h	M																			
	C												5		5	5				
a	M								22						22	22				
	C								1						1	1				
Totals		39	77	58	2	42	22	240	89	144	5	121	25	37	421	661	106	88	194	6

TABLE 6.7.8

SUMMARY OF CODED BEHAVIOUR : TEACHER 6

		Teacher-individual							Teacher-class								Student			
		1	2	3	4	5	6	Tot	1	2	3	4	5	6	Tot	Tot	7	8	Tot	X
+	M			12				12								12				
-	M										2				2	2				
1	M	4	20	32	4	37	21	118	82	46	5	16	18	60	227	345	39	27	66	
	C	42	20		4	34	4	104		54		74	49	2	179	283	37	19	56	
0	M																			
	C																			
h	M																			
	C					3		3					21		21	24				
a	M																			
	C																			
Totals		46	40	44	8	74	25	237	82	100	7	90	58	62	429	666	76	46	122	16

TABLE 6.7.9

SUMMARY OF CODED BEHAVIOUR : TEACHER 7

		Teacher-individual							Teacher-class								Student			
		1	2	3	4	5	6	Tot	1	2	3	4	5	6	Tot	Tot	7	8	Tot	X
+	M			3				3			3				3	3	3		3	
-	M			1				1			1			1	2	3				
l	M	5	27	48	5	61	48	194	40	16	11	8	11	64	150	344	68	64	132	
	C	9	64	22	10	65	14	184	1	27	2	52	18	1	101	285	105	88	193	
o	M					1		1								1				
	C	3				8		11					8		8	19		2	2	
h	M																			
	C																			
a	M																			
	C																			
Totals		17	91	74	15	135	62	394	41	43	17	60	37	66	264	658	176	154	330	14

TABLE 6.7.10

SUMMARY OF CODED BEHAVIOUR : TEACHER 8

		Teacher-individual							Teacher-class								Student			
		1	2	3	4	5	6	Tot	1	2	3	4	5	6	Tot	Tot	7	8	Tot	X
+	M			4				4			1				1	5	11		11	
-	M																			
i	M	6	26	28	3	39	29	131	47	20	8	5	10	44	134	265	64	44	108	
	C	63	29	3	15	37	1	148	2	27		21	20		70	218	67	66	133	
o	M					1		1								1				
	C	7	1			11		19								19				
h	M												1		1	1				
	C												7		7	7				
a	M																			
	C																			
Totals		76	56	35	18	88	30	303	49	47	9	26	38	44	213	516	142	110	252	15

6.8 VALIDITY AND RELIABILITY

Validity

Nuthall and Church (1972) and Rosenshine and Furst (1973) observe that researchers typically refer to indices of rater agreement or coder agreement as estimates of the reliability of classroom observation systems but, just as typically, neglect consideration of the validity of the system. In talking about the reliability or the validity of an observation system, Rowley's (1975) point is well taken: reliability is a property of a measure, not of an instrument, or a record. So the phrase "the reliability of a system" is taken here to mean "the reliability of measures resulting from the use of the system in its specified way". Rosenshine and Furst believe that the author of an observational instrument should first attempt to determine the predictive validity of the variables. In a one-off study the relationship between predictor and criterion variables is usually the end point of the study. The difficulty lies in finding an established system (for which data on validity and reliability may be available) that is entirely appropriate to a particular investigation. As Nuthall and Church point out it is likely that an investigator will have to develop his own system to meet the particular needs of his research (1972, p.504). In any case, as mentioned earlier, Nuthall and Church caution against the selection of an observation system on such criteria because "we do not know enough about the causes and effects of teacher behaviours to be able to lay down the standards which observation systems must meet if they are to be considered valid and reliable instruments" (ibid.).

Medley and Mitzel (1963) are more concerned with aspects of content validity than criterion validity. They contend that the validity of a measurement of classroom behaviour depends on three conditions: "(1) A representative sample of the behaviors to be measured must be observed. (2) An accurate record of the observed behaviors must be obtained. (3) The records must be scored so as to faithfully reflect differences in behavior" (p.250). These three criteria are now discussed in a different order.

In the previous section procedures are described by which it was hoped to secure, within the limits of practicability, an accurate record of the teachers' behaviour. As observed earlier, it is undoubtedly true that not all the behaviour that occurred in the classroom was recorded. Some information is lost in the process of recording audiotapes, but the record of the teachers' verbal behaviour is virtually complete. Since the recording of non-verbal behaviour depends on the perception of the observer and his ability to capture, in longhand, the subtleties of facial expression, bodily posture and movement, it is clear that this record will be far less complete. However, by focussing on contradictory non-verbal behaviour and broad descriptions of the teachers' activity during periods of silence, it is believed a useful dimension was added to the audiotape record.

Whether or not these records are scored so as to faithfully reflect differences in behaviour depends on the characteristics of the category system and the coders. In the development of a category system an attempt is made to define the various categories so that the coder is required to make few if any inferences about, or interpretations of, the behaviour to be categorized. In the development of CATS a further attempt was made to describe the categories in such a way that a user of the instrument did not require lengthy training to develop consistent patterns of scoring. It would, for instance, be possible to develop a highly idiosyncratic system which required that all teacher grimaces be recorded as 'positive affect' and by lengthy training ensure that observers reached a high degree of agreement. In this way high reliability could be achieved at the expense of validity. However, in this study coders did not receive special training in the use of the category system: they were merely asked to code the annotated transcripts according to the procedures and definitions set out in the manual for CATS (Appendix II). Whether the category system is in itself sufficiently extensive must be judged by inspection and by seeing how many behaviours are classified as non-categorizable. It is contended that, in this situation, substantial agreement between different workers coding the same transcript, and substantial consistency by the same worker coding the same transcript on different occasions, are indices not only of aspects of the reliability of the system but of face validity, and could be regarded as tests of Medley and Mitzel's third criterion.

To observe and record a sample of teacher behaviours that could be regarded as representative is a more difficult problem. Rosenshine (1973) reanalyzed some data from a study by Moon (1971) and showed that, when using a five-category system for classifying teachers' questions (Fischler & Anastasiow, 1965), teachers varied considerably from lesson to lesson and that different teachers varied in different ways. Using the Spearman-Brown prophecy formula he estimated that, for the most stable category, 11-20 observations would have been needed to obtain a stable mean for each teacher. Rosenshine concluded that "the few studies which have been conducted to date suggest that for many of the cognitive variables, whether teachers are trained or untrained, even 30 observations will be insufficient to obtain a trustworthy mean score" (p.226). However, for affective variables Rosenshine adds that one or two observations appear to be sufficient to give stable mean scores for individual teachers. The difference in stability between cognitive and affective behaviours is consistent with the distinction drawn earlier in this chapter between strategy and style. Rowley (1976) presents less pessimistic indications. In estimating reliability by analysis of variance, and using the modified version of Herbert's (1967) System for Analyzing Lessons, he found that "reliabilities of measures based on single visits (r_{11}) are all quite low (below 0.4, in all except three cases); but for some categories, the reliabilities of average measures based typically on 6-7 visits (r_{nn}) are substantial, exceeding 0.8 in some cases" (p.56). Like Rosenshine, he notes "the relatively high reliabilities recorded for most of the categories of the dimension 'Affect Tone'" (ibid.).

Although the Web of Life course has a grand design with respect to anticipated outcomes, the teaching strategies recommended to achieve these outcomes vary with the topic. Some sections of the course call for class discussion while other sections call for individual laboratory investigations to be carried out by the students. Care was taken to choose five lessons for each teacher over the school year which sampled the range of teaching strategies suggested in the curriculum materials. In terms of both content and general strategy a similar sample of lessons was taken from each teacher. Because the lesson time-table and syllabus was known for each teacher it was possible for the investigator to select lessons in advance and to arrive at each school without prior notice so that no special preparation would be made. Such a sample may not provide sufficiently trustworthy mean scores for individual teachers on cognitive aspects of classroom interaction, although they may be somewhat better for managerial and interpersonal aspects of teaching style. However, Rosenshine and Furst (1973) observe that if the purpose of the study is to compare the behaviours of one group of teachers with another group "then the number of observations necessary to obtain a representative group mean may be small" (p.169). So that for the present research the number of observations should be adequate for the two main purposes of the category system: to establish that there are two groups of teachers showing contrasting teaching styles, and to describe these teaching styles in terms of the group means.

Reliability

Reliability can be considered as the consistency with which an instrument measures whatever it does measure, whereas validity is the extent to which the instrument measures what it is intended to measure. Lack of consistency in the data generated by a category system can result from three major causes:

(a) inconsistency on the part of the coder,
(b) lack of agreement between coders, and
(c) lack of stability in the teachers' behaviour over time.

Validity is the prime consideration, but is difficult to estimate. However, reliability sets a ceiling on validity, and it is important to know the source and magnitude of errors before attempting to relate independent and dependent variables. Estimates of reliability are reported by a number of correlation indices, depending on the nature of the data and the relationships of the two observations to one another. Such indices appropriate to nominal scale coding are now discussed.

An index for nominal scale coding. Of the indices of agreement for nominal scale coding discussed by Emmer (1972), the statistic Kappa (κ) described by Cohen (1960) is preferred to percentage agreement and Scott's (1955) Pi (π) coefficient.

The percentage agreement is biased in favour of systems with few categories because of the increased proportional frequency of chance agreement. Bennett, Alpert, and Goldstein (1954) have proposed a correction for the number of categories:

$$S \text{ (the index of consistency) } = \frac{k}{k-1}\left(P_o - \frac{1}{k}\right)$$

where P_o is the observed percentage agreement between coders and k is the number of categories. It is assumed that all categories have an equal probability of use, and so the index will be inflated by the inclusion of categories that are little used.

Scott's π is given by:

$$\pi = \frac{P_o - P_e}{1 - P_e}$$

where P_o is the observed percentage agreement between coders and P_e is the percentage agreement that could be expected by chance. It is assumed that the distribution of proportions over the categories is known for the population and is taken to be equal for the coders. P_e is then calculated by summing the squares of these proportions over all categories. Cohen (1960) observes that the assumption of equal distribution by coders over the categories is questionable "since one source of disagreement between a pair of judges is precisely their proclivity to distribute their judgements differently over the categories" (p.41).

The coefficient κ is given effectively by the same formula as π:

$$\kappa = \frac{P_o - P_c}{1 - P_c}$$

where P_o is the proportion of units in which the judges agreed, and P_c is the proportion of units for which agreement is expected by chance. However, P_c is defined differently from P_e in Scott's formula.

In Cohen's definition, P_c is the sum of the joint probabilities of the marginals, so that no assumption need be made about the identity of judges' distribution over the categories.

Coder consistency. It is important to know how consistent a coder is in coding identical audiotapes and transcripts on different occasions. Lack of consistency by a coder will set ceilings on all other estimates based on the coded data.

At the end of the initial coding phase the two coders were asked to recode a sample of eight lessons, one from each teacher, from the 40 originally coded. The mean values of κ for the two coders, as an estimate of their consistency, were 0.80, and 0.76. Given minimal training with the category system, these values appear to reflect an acceptable level of stability on the part of the coders, and their acceptance of the category definitions, which could undoubtedly be improved with additional training.

Inter-coder agreement. As observed by numerous writers (e.g., Medley & Mitzel, 1958; Weick, 1968; McGaw, Wardrop, & Bunda, 1972; Rosenshine & Furst, 1973) most observational studies report reliability almost exclusively in terms of observer agreement. In this study the equivalent index is one of agreement between coders, since coders are working from audiotape records and annotated transcripts rather than observing the lesson directly. Although it is clear that it is necessary to obtain substantial agreement between coders, the conditions under which this agreement is obtained are of some importance. It is possible for an investigator to improve inter-coder agreement by one of two methods: either by defining the categories more clearly, and removing ambiguities, or by giving more extensive training to the coders. As observed earlier, inflated indices of inter-coder agreement can be produced, for even the most idiosyncratic system, by a sufficiently intensive training program. In the development of CATS it was decided not to train coders, but to rely entirely on the procedures and definitions set out in the manual. In these circumstances an index of inter-coder agreement, as well as reflecting on the objectivity of the system, reflects also on the acceptability of the category definitions as a basis for classifying natural behaviour and so tells something about the validity of the instrument. However, this index must also reflect error due to rater bias and this is an aspect of reliability.

The task for the coders is to categorize units of behaviour into one of 77 possible categories. Some of the categories can be said to be related to one another in a qualitative way, but others are quite dissimilar. The range of categories must then be considered as a nominal scale, and interval-scale or ordinal-scale statistics would be inappropriate to measure the extent to which two coders (or the one coder on different occasions) agree on the allocation of a collection of units of behaviour to the different categories. Eight lessons were chosen at random, one from each teacher, as a basis for calculating the agreement between the two coders, independently coding identical audiotapes and annotated transcripts. The values of κ are shown in Table 6.8.1.

TABLE 6.8.1

ESTIMATES OF INTER-CODER AGREEMENT

Lesson	14	24	33	45	53	66	73	84	M	SD
κ	0.72	0.63	0.79	0.84	0.70	0.80	0.72	0.68	0.74	0.07

It is difficult to know what an acceptable level of intercoder agreement should be. Smith and Meux (Hyman, 1968) say "coefficients [of agreement between judges] range from 0.00 to 1.00. The median is 0.67, and the middle 50% of the coefficients range from 0.62 to 0.84 - a fairly high percentage of agreement for the present status of the categories" (p.114). Withall (1949) reports mean percentage agreements between judges and the researcher of 64%, 66%, 56%, and 75% for his Climate Index. On the other hand Flanders (1967) says that "a Scott coefficient of 0.85 or higher is a reasonable performance" (p.166). Parakh (1968) reports that "Scott's coefficients of 0.8 to 0.9 were usually obtained" (p.289, emphasis mine). Evans (1969) does even better with Scott's coefficients of 0.92, 0.95 and 0.93 for inter-observer agreement. Parakh does not comment on the length of training for his observers, but it is clear that Evans' period of training was quite extensive: "no time limit was placed on the initial observing, classifying and recording of behaviors" (1969, p.222). The coefficients in Table 6.8.1 could be seen to be encouraging in the context of Smith and Meux, and Withall, but disappointingly low in the context of Parakh, and Evans. An inspection of the coding sheets of both workers indicates that most disagreement occurred in the coding of the development and non-verbal categories. Both coders admitted difficulty in distinguishing clearly between presentation of new information and the development of old. The researcher, being thoroughly familiar with the Web of Life course and having carried out the original recording of the lessons, appears likely to be in a better position to distinguish between the two than the second coder, who was working from the tapes and transcripts alone. Similar problems were associated with the coding of non-verbal behaviour from annotations in the transcript and associated auditory clues from the audiotape. It is clear that, in these categories of the system, it will be more difficult to secure agreement, but it is felt that the values of κ presented in Table 6.8.1 represent an acceptable measure of agreement, and reflect reasonably well on the adequacy of the procedures and definitions for their intended purpose.

Stability of Teachers' Behaviours. Medley and Mitzel (1963), Rosenshine (1970 , 1973), McGaw, Wardrop, and Bunda (1972), Cronbach, Gleser, Nanda, and Rajaratnam (1972), have all clearly demonstrated the limitations of using inter-coder or inter-observer agreement as the sole index of reliability.

> The confusion introduced into the literature through failure clearly to distinguish the different sources of unreliability, and through over-emphasis on inter-judge agreement has resulted from a confusion of the importance of primacy with prime importance. Inter-judge agreement is the first but not the most important issue to be faced. (McGaw et al., 1972, p.16)

The stability of measures of teacher behaviour has not been adequately examined or established in most classroom research. If a researcher is attempting to determine the relationship between observed classroom behaviour and some outcome variable it is important to know that the behaviour measured is stable. However, in considering the temporal instability of teacher behaviour it must be appreciated that teachers consciously adapt their behaviour to different teaching situations. The concept of stability proposed by Medley and Mitzel (1963) does not allow for such deliberate adaptation. McGaw, Wardrop, and Bunda (1972) observe that Medley and Mitzel's requirement of stability "may be a reasonable assumption in relation to relatively enduring aspects of personality" but unreasonable when considering "lawful adaptations of behavior to different situations" (p.16). Flanders and Simon (1969) refer to the concept of "flexibility of teacher influence" (p.1428). Changes of behaviour to accommodate dissimilar classroom activities such as individual laboratory work and class discussion would appear to be lawful. Such lawful changes in behaviour are, in this investigation, regarded as changes in strategy. The Teacher's Guides, (AAS, 1969, 1976) give advice and suggestion on just this kind of adaptation.

However useful the distinction between strategy and style may be, it is quite clear that they can affect one another. For example, one teacher may be stylistically characterised as "running a tight ship", but his managerial ploys may not be nearly as evident in a class discussion, guided by the teacher, as when students are moving about in laboratory work. Strategy can affect style. Similarly, Tamir (1975) presents evidence of teachers' 'curricular bias', some being strong supporters of the inquiry-centred strategies of BSCS biology courses and others who exhibit a less positive approach to these strategies. When strategy and style are dissonant it is likely that strategy will be modified. For example, it would be possible to attempt to teach inquiry processes didactically: "Children, take these notes. Scientific inquiry takes place in the following way ...." For this reason, a sampling of lessons was sought that would reflect the range of teaching strategies recommended in the Web of Life curriculum materials. The stability of measures obtained by CATS will depend to a large extent on the extent to which CATS detects those "relatively enduring aspects of personality", and habitual patterns of personal interaction; that is, the extent to which it measures style rather than strategy.

Some early studies tended to emphasize the general stability of teacher behaviour (Hughes, 1962; Bellack et al., 1966; Gallagher, 1967) but the more recent studies of Moon (1971), Emmer and Peck

(1973) and Rosenshine (1973) indicate considerable variability of teacher behaviour over time. Campbell (1972) showed that some teachers are consistent over time while others vary considerably. Shavelson and Dempsey's (1976) review examines a number of studies which provide data on the stability of teacher behaviour by clusters of categories. They conclude that the findings on the stability of most measures of teacher behaviour are equivocal. Median coefficients of stability over the various categories for the studies (cited in Shavelson & Dempsey, 1976) which used frequencies of behaviour as their units of measurement are: Friedman, 0.58; Moon, 0.42; Trinchero, 0.31; Ehman, 0.35; and Bellack, 0.66.

The question of the stability of teacher behaviour is an empirical matter, and in the present investigation, has been determined in the following manner. The frequencies of behaviour in each category for each recorded lesson were ranked. These ranks could then be correlated with the ranks of the total frequencies (over five lessons) for that teacher. However, the correlation would be spuriously inflated if the total included the frequencies of the one lesson being considered, particularly since there are only five lessons in each total. So correlations were calculated between ranks for each lesson and ranks from a composite of the remaining lessons for that teacher, providing an unbiased rank order correlation (cf. De Landsheere, cited in Rosenshine, 1973, p.211). These coefficients indicate the degree to which the total pattern of behaviour in a particular lesson reflects the aggregate pattern of behaviour for that teacher, and are given in Table 6.8.2.

TABLE 6.8.2

ESTIMATES OF STABILITY OF TEACHER BEHAVIOUR[a]

Teacher	Lesson					Mdn
	1	2	3	4	5	
1	0.58	0.76	0.77	0.71	0.43	0.71
2	0.63	0.73	0.72	0.64	0.66	0.66
3	0.55	0.55	0.78	0.84	0.49	0.55
4	0.[illegible]	0.72	0.52	0.74	0.65	0.72
5	0.77	0.74	0.76	0.77	0.72	0.76
6	0.55	0.34	0.74	0.53	0.63	0.63
7	0.57	0.59	0.76	0.32	0.67	0.67
8	0.75	0.84	0.87	0.78	0.80	0.80

[a] Unbiased Spearman's rank-order correlation coefficients

These data suggest a number of interesting lines for further investigation. In what ways are some teachers more consistent in their behaviour than others? What happened in lessons 6 2 and 7 4 that their coefficients are much lower than the others? However intriguing, questions such as these are considered to lie outside the scope of the present study.

In view of the pessimism of Rosenshine (1973), the values of stability coefficients from other studies, and the observation by Shavelson and Dempsey (1976) that "frequency counts tended to have the lowest bound stability coefficients" (p.608), these estimates can be taken as encouraging. They may, of course, be interpreted in a number of ways. However, they should not be taken to be representative of all teachers, nor of all biology teachers: the teachers in this investigation were selected to provide contrasting teaching styles. (Anticipating the validation of the treatment groups it is interesting to note that the median coefficient for one group is 0.74 and for the other, 0.73. Stability of behaviour then is not a distinguishing characteristic between the groups.) Within the context of this investigation, these estimates are regarded as conservative, and as estimates of the stability of style they may be to some extent influenced by the range of strategies employed by teachers in different teaching contexts. Considering this possibility, the values in Table 6.8.2 give grounds for believing that a sufficiently representative sample of lessons was recorded to provide measures of teacher behaviour that are stable enough for group comparisons.

6.9 VALIDATION AND DESCRIPTION OF TREATMENTS

To summarize the position so far. A number of judges were asked independently to provide two lists of biology teachers, whose teaching was known to them, who fitted one of two dissimilar descriptive prototypes. Not all teachers were known to all the judges, but where a teacher's name appeared as having been categorized in the same fashion by at least three judges this was taken to be substantial evidence of the impression created by this teacher's classroom behaviour. Four teachers were finally chosen from each of these contrasting lists to take part in the present study. However, the investigator was unwilling to categorize the teachers on this basis; that is, to assert that they behaved as they were reputed to behave. It was merely a device to secure two groups of teachers who could reasonably be expected to exhibit contrasting teaching behaviour. In order to gather more detailed and objective data on the classroom behaviour of these teachers the CATS system was developed, and some attention has been paid to the validity and reliability of such data.

Arguments have been advanced above that the content, objectives, and teaching strategies have been similar for all teachers, and that any remaining differences will be due to enduring patterns of behaviour developed through training and teaching experience and those related to more fundamental personality characteristics. In other words any similarities and contrasts can be ascribed to differences in teaching style. On the basis of the data gathered, how may the eight teachers be grouped on the basis of similarity or homogeneity of teaching style? As Tatsuoka (1973) points out, one approach to this problem is to use an 'inverted' factor analysis, often referred to as the Q-technique, pioneered by Stephenson (1936, 1953). Such an analysis begins with an intercorrelation matrix based on pairs of individuals rather than on scores from pairs of tests. Such a procedure will give factors which can only "represent linear configurations in the person space, whereas the general topography of types or clusters can be more complex" (Tatsuoka, 1973, p.283). And, as correlation coefficients do not reflect the differences that may exist in elevation of the two score profiles being correlated, the factor analysis, similarly, will not reflect such differences. These difficulties have led to the exploration of fundamentally different approaches to solving the problem of clustering. Tatsuoka (1973) discusses the difficulties of Ward's (1963) hierarchical grouping system, and Cattell and Coulter's (1966) production of the Taxonome computer program. White (1973) discusses differences in measures of similarity adopted by the Q-technique, which is based on correlations, and those adopted by Taxonome and H group (Veldman, 1967), which are based on Pythagorean distance between scores. White further points to differences in the meaning of 'similarity' and to the method of forming groups, particularly as they affect H group and Taxonome.

The use of any one method of clustering involves a number of subjective decisions. To provide some check on the validity of the clustering, it was decided to subject the data on teacher behaviour, presented in Tables 6.8.3 to 6.8.10, to analysis by an inverted factor method, and the H group. For the inverted factor analysis, the minimum residuals method (Comrey & Ahumada, 1964) was used with rotation of factors by the normal varimax method (Kaiser, 1958). Figure 6.9.1 shows teachers plotted according to their rotated factor loadings.

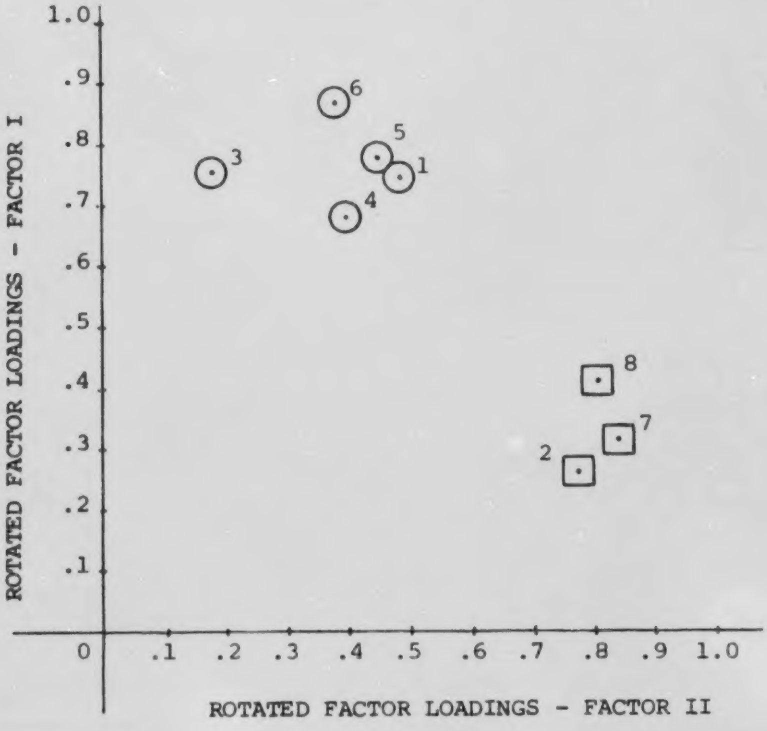


Figure 6.9.1. Clustering of teachers by inverted factor analysis

The inverted factor analysis has clustered teachers into two groups on the similarity of their profiles on 45 variables of classroom behaviour, although within a group the height of the profile for individuals may vary. Using the same data on the 45 variables of classroom behaviour H group (which is based on different measures of similarity, and on different assumptions about orthogonality) gives the same solution, grouping teachers 2, 7 and 8 together and grouping teachers 1, 3, 4, 5 and 6 together, so that great confidence can be placed in the validity of the result.

Figure 6.9.1 gives a visual representation of the remarkable degree to which the teachers do cluster. The clarity of the separation into two groups is striking, although the clustering is not what was expected. It was expected that teachers 1, 3, 4, and 5 would be similar to one another in teaching style, but contrasting with 2, 6, 7, and 8 who were also expected to be similar to one another. It is quite clear from conversation with two of the judges that they believe teacher 6 was correctly categorized initially, and that he has a reputation as being a flexible, process-centred teacher. However, sustained observation and analysis of his classroom behaviour places him in the other group. The record of classroom behaviour is not influenced by a teacher's reputation, which may be developed largely through out-of-class involvement with other teachers in matters of professional concern. Medley and Mitzel (1963) point to the importance of increasing the number of observations, rather than observers, in securing reliable measures of classroom behaviour. And although the judges may have known the teachers for a number of years, it is unlikely that they would have observed as much of their classroom behaviour as was observed in this study. Quite apart from considerations of reliability, it is believed that the systematic observation and analysis of classroom behaviour has provided a more valid basis for categorizing teachers into contrasting treatment groups.

It has been established that the teachers in this investigation fall into two contrasting groups in terms of teaching style, and are similar in style within each group. How can the contrasting styles now be described? Rosenshine's (1973) study suggests that because of the instability of teacher behaviour it is unwise to take just a few observations of a teacher's behaviour as a representative sample. However, for the purpose of comparing the behaviour of one group of teachers with another group, Rosenshine suggests that "one or two observations are sufficient to obtain a trustworthy mean score for the group" (p.226). In Tables 6.9.2 and 6.9.3 the mean frequencies of behaviour for the two groups of teachers are given. The frequencies relate to a one-sixth sample of 250 recorded minutes as outlined earlier in this chapter. In Table 6.9.2 the units of behaviour are compared by gross categorical distinctions, and in Table 6.9.3 by the finer distinctions made in the category system employed.

The first level categories in Table 6.9.2 show similar profiles of behaviour between the groups with the exception of the level of pupil response. The group of teachers 2, 7, and 8 will now be referred to as Group 1, and teachers 1, 3, 4, 5 and 6 as Group 2. Although they were originally selected to conform to the descriptive prototype A, they are not now the same group, and their classroom behaviour must be described in terms of observed behaviour (the CATS data), rather than expected behaviour (the prototypes A and B). Both groups show higher frequencies of what might be called indirect (Amidon & Flanders, 1961; Flanders, 1965) influence (development, evaluation, questioning), rather than direct influence (presentation, direction). This possibly reflects the acceptance, by both groups of teachers, of the indirect inquiry-oriented strategies recommended in the curriculum materials.

The second level categories show the same overwhelming preponderance of closed behaviour for both groups. The higher levels of open behaviour for Group 1 are almost entirely due to the open questions asked by this group, in comparison with the almost total absence in Group 2. There is, for both groups, a heavy emphasis on convergent memory questions. For both groups there are low frequencies of affective evaluations, with positive reinforcement three times as frequent as negative reinforcement. The frequency of use of audio-visual media, mostly the chalkboard, but less frequently biological equipment and materials, was similar for both groups.

The most striking difference in observed classroom behaviour between the groups lay in their preferred teaching modes. Group 1 carries out almost twice the number of teaching acts with individuals, or small groups of students, than with the class as a whole: for Group 2 this ratio is reversed. Inspection of Table 6.9.3 reveals the finer detail. In presenting new course content Group 1 converses with individuals and small groups significantly more often than Group 2 (variable 12), and the converse is also true: Group 2 lectures to the class as a whole significantly more often than Group 1 (variable 31). Similarly, Group 1 asks questions, both closed (memory) and open questions, of individuals (variables 14, 15, 16) more frequently than Group 2. The general picture is of Group 1 teachers, in comparison with Group 2, spending less time lecturing from the front of the class, and more time moving around the classroom giving assistance to individuals at their desks, or small groups working on laboratory investigations. This picture is supported by

TABLE 6.9.2

SIGNIFICANCE TESTS FOR DIFFERENCES BETWEEN INSTRUCTIONAL TREATMENTS BY GROSS CATEGORICAL DISTINCTIONS

Category	M_1 (Teachers 2,7,8)	SD_1	M_2 (Teachers 1,3,4,5,6)	SD_2	t
First level categories					
1. Non-verbal	91.00	33.51	128.20	41.05	-1.316
2. Development	159.67	72.23	176.40	75.91	-0.307
3. Evaluation	76.67	28.29	60.00	18.82	1.018
4. Presentation	75.00	31.00	88.60	24.57	-0.693
5. Questioning	168.00	40.15	112.20	43.87	1.791
6. Direction	92.33	30.89	77.40	26.31	0.732
7. Pupil response	152.33	20.55	88.60	33.73	2.910*
8. Pupil initiation	108.33	46.52	69.80	32.76	1.392
Second level categories					
Open behaviour	13.33	11.55	0.20	0.45	2.693*
Closed behaviour[a]	628.67	145.50	609.40	59.17	0.180
Positive affect	9.00	6.08	7.00	3.81	0.584
Negative affect	3.00	3.00	2.00	2.45	0.518
Memory questions	150.67	42.67	101.80	37.79	1.695
Higher-order questions	7.67	7.51	9.40	8.96	-0.279
Open questions	9.67	8.74	0.20	0.45	2.563*
Audio-visual	109.33	35.85	119.04	39.58	-0.310
Third level subscript					
Individual mode	427.67	143.15	219.20	50.00	3.097*
Class mode	235.00	26.91	423.60	55.68	-5.375**
Fourth level categories					
Teacher Managerial	352.00	79.02	386.60	32.16	-0.900
Teacher Substantive	309.67	68.68	252.20	67.42	1.160
Pupil Managerial	118.33	17.01	93.20	36.30	1.102
Pupil Substantive	142.33	48.68	63.80	35.91	2.648*

*$p < 0.05$; **$p < 0.01$

Critical t = 2.447 at 0.05 level of confidence

[a] Includes neutral evaluative statements, and memory questions

TABLE 6.9.3

SIGNIFICANCE TESTS FOR DIFFERENCES BETWEEN INSTRUCTIONAL TREATMENTS BY FINE CATEGORICAL DISTINCTIONS

Variable		M_1 (Teachers 2,7,8)	SD_1	M_2 (Teachers 1,3,4,5,6)	SD_2	t
1.	(10$_i$C)	3.33	3.51	0	0	2.251
2.	(11$_i$M)	4.33	2.08	4.00	3.54	0.146
3.	(11$_i$C)	43.67	30.09	25.00	11.92	1.284
4.	(20$_i$C)	0.33	0.58	0	0	1.369
5.	(21$_i$M)	44.00	30.32	31.60	19.93	0.710
6.	(21$_i$C)	66.00	38.04	25.40	25.32	1.843
7.	(31$_i$M)	48.67	21.01	43.80	13.94	0.401
8.	(31$_i$C)	9.00	11.27	0.20	0.45	1.849
9.	(3+$_i$M)	7.33	6.66	6.66	3.78	0.204
10.	(3−$_i$M)	1.33	1.53	1.60	2.61	-0.158
11.	(41$_i$M)	6.00	3.61	2.00	1.41	2.301
12.	(41$_i$C)	18.00	9.85	1.80	1.30	3.835**
13.	(50$_i$M)	0.67	0.58	0.20	0.45	1.292
14.	(50$_i$C)	6.33	5.69	0.00	0	2.642*
15.	(51$_i$M)	59.00	19.08	31.40	5.55	3.173*
16.	(51$_i$C)	62.67	24.58	17.60	10.43	3.728**
17.	(5h$_i$C)	0	0	1.40	1.34	-1.750
18.	(61$_i$M)	39.67	9.71	16.40	9.92	3.235*
19.	(61$_i$C)	6.67	6.66	2.60	2.41	1.290
20.	(11M)	37.67	10.69	85.00	34.18	-2.268
21.	(11C)	1.00	1.00	3.20	6.61	-0.555
22.	(1aM)	1.00	1.73	9.60	12.74	-1.127
23.	(1aC)	0	0	0.20	0.45	-0.750
24.	(21M)	21.67	6.66	55.20	24.07	-2.293
25.	(21C)	27.33	0.58	62.00	35.92	-1.618
26.	(31M)	6.67	5.13	5.80	3.70	0.280
27.	(31C)	1.00	1.00	0.80	1.30	0.226
28.	(3+M)	1.67	1.56	0.40	0.55	2.161
29.	(3-M)	0.33	0.58	0.40	0.89	-0.114
30.	(41M)	15.00	14.80	20.00	15.57	-0.447
31.	(41C)	36.00	15.52	64.60	14.22	-2.670*
32.	(50C)	2.67	4.62	0	0	1.369
33.	(51M)	11.33	1.53	15.40	8.96	-0.756
34.	(51C)	17.67	2.52	35.80	26.78	-1.133
35.	(5hM)	0.33	0.58	0	0	1.369
36.	(5hC)	7.33	7.51	8.00	7.81	-0.118
37.	(61M)	45.00	18.52	45.80	11.50	-0.077
38.	(61C)	0.67	0.58	3.60	3.65	-1.341
39.	(6-M)	0.33	0.58	0	0	1.369
40.	(71M)	64.00	9.64	47.40	17.98	1.448
41.	(71C)	79.33	22.23	36.80	21.18	2.704*
42.	(7+M)	5.33	4.93	3.60	3.29	0.607
43.	(80C)	2.67	4.62	0	0	1.369
44.	(81M)	45.33	18.04	42.20	17.12	0.246
45.	(81C)	62.33	27.68	27.00	16.02	2.343

*$p < 0.05$. **$p < 0.01$

Critical t = 2.447 at 0.05 level of confidence other differences which almost reach the 0.05 level of significance: greater frequencies of variable 1 (non-verbal, open, content-centred behaviour, e.g., assistance) given to individuals by Group 1; greater frequencies of variable 20 (non-verbal, closed, management-centred behaviour, e.g., supervision) exercised over the class by Group 2; and greater frequencies of variable 24 (closed, management-centred development such as explanations or reiterations of procedural rules) by Group 2.

The fourth level categories show a higher incidence in both groups of managerial behaviour than content-centred behaviour. This reflects to some extent the substantial frequencies of procedural directions given associated with individual laboratory work, and the time spent in exercising supervision by silently walking around the laboratory (although with a greater readiness to give assistance by Group 1 than Group 2). Insofar as pupil behaviour reflects, to some extent, the classroom conditions and expectations generated by the teacher, it is interesting to note the significantly higher frequencies of pupil response associated with content development in Group 1 than in Group 2. This content-oriented participation by pupils in Group 1 is further supported by higher frequencies of content-centred pupil initiation (variable 45) which almost reach the 0.05 level of significance.

Class size may have a considerable bearing on the ways in which a teacher may behave in the classroom. For example, it may be argued that a class-centred teaching style is likely to be more common in large classes than in small ones. In comparing the class sizes of the two groups for this purpose it is appropriate to include repeat students, those who were absent from the final testing, and those who withdrew during the year. Including these students, the class sizes for Group 1 were 16, 25, and 30 (mean 20.2) and for Group 2 were 25, 18, 19, 24, and 15 (mean 23.7). The difference between the means is not statistically significant (t = 0.64; critical t(6) = 2.45 at the 0.05 level of significance) and so in this particular investigation class size can not be regarded as contributing significantly to treatment differences. It is also interesting to note that in the largest class the individual-centred teaching style was evident.

The mere counting of acts of classroom behaviour does not of course adequately describe teaching, nor does the comparison of frequencies of particular kinds of teaching behaviour fully describe the differences between teaching styles. There is in this investigation no analysis of the classroom interaction, no analysis of the sequence of behaviours: what happens after a teacher gives a direction? how does a teacher deal with a student's contribution? There is no analysis of the quality of the statements made: how accurate is the information presented? how clear are the explanations? how well-suited is the explanation to the student's present level of understanding? The act of teaching, like all interpersonal interactions, is extraordinarily complex and difficult to summarize. A loss of data always accompanies summary. In this investigation it was decided to record teachers' classroom behaviour and to tally frequencies of particular categories of behaviour for two principal purposes: to check on the validity of the assumption of two treatment groups, and to provide some basis for describing these treatments. The data have been well-suited to the first purpose, not nearly so well-suited to the second. It would, indeed, be possible to analyze the audiotapes and transcripts paying attention to sequence and to quality, but it was considered to be outside the scope of the present investigation. What in fact we have is something like a black and white picture of teaching: it could have been in colour, it could have been a motion picture, and it could have been both.

CHAPTER 7

THE CRITERION TESTS

7.1 INTRODUCTION

7.2 MEASURES OF COGNITIVE OUTCOMES

  Construction of achievement tests

  Sampling of process objectives

  Estimates of reliability

7.3 MEASURES OF NON-COGNITIVE OBJECTIVES

  The role of the laboratory

  The measurement of attitudes

7.4 THE BIOLOGY ATTITUDE INVENTORY SCALES

  Description

  Estimates of reliability

7.1 INTRODUCTION

It is often difficult to make valid comparisons of contrasting instructional treatments since they tend to rest on different assumptions and to be directed towards different outcomes. It is suspected that the inconclusive nature of the research on contrasting instructional treatments may be partly due to the fact that the criterion tests employed are likely to favour one program over the other. For example, where traditional content-orientated achievement tests are used, it would not be surprising to find that traditional content-orientated methods of teaching lead to greater achievement than do process-orientated methods.

In the present investigation all teachers are teaching the same course towards the same agreed ends using the same instructional materials and the same recommended teaching strategies. The differences in instructional treatment between the groups lie in the teachers' own natural teaching styles. In this situation criterion tests designed to measure the agreed outcomes of the program should favour neither group.

In the Teacher's Guide (AAS, 1973), four broad course objectives are set out:

> The overall aims are to foster in the individual student: 1. Scientific understanding of the living world, including man. 2. Understanding of the nature, scope and limitations of science. 3. Competence and willingness to apply biological understanding to scientific and everyday problems, and to approach these problems in scientific ways. 4. Ability to communicate effectively, and to work jointly with others towards solving problems of mutual concern. (pp.6-7)

The authors of the Guide then discuss these objectives in some detail and recognize that the non-cognitive outcomes implicit in them are not likely to be tested by conventional written tests such as used in public examinations. This issue will be discussed in section 7.3.

7.2 MEASURES OF COGNITIVE OUTCOMES

Construction of achievement tests

The content goals of the course have been divided into a number of 'major ideas' or major generalizations of biological science. Activities are set out for the students which it is hoped will lead them to form these generalizations for themselves. Each major idea is built up from a number of individual concepts or 'single propositions'. For example:

> Organisms of different species occur together in nature as communities.
> Organisms with similar requirements, and those that depend on each other, may be found living together.
> Groups of organisms, called communities, can be identified.
> In nature there is a continuity in space between one community and another.
> Delimiting communities is as subjective as are other classifications. (Teacher's Guide, AAS, 1973, pp.64-5)

These single propositions are arranged in appropriate teaching sequences and indicate the desired outcomes of the various student activities.

The authors of the Web of Life course have developed a hierarchy of process goals based on Bloom's Taxonomy (Bloom, et al., 1956). It is expected that

> the student should, for each concept developed - whether related to the living world or a scientific process component - show some ability to:
>
> 1. (a) State, explain, interpret and exemplify the concept;
>   (b) Use, in their generally accepted senses, fundamental terms and classifications related to the concept;
>   (c) Cite, and explain or interpret, scientific evidence in support of the concept.
>   (d) For concepts related to the living world, state and explain how evidence is collected; and
>   (e) How the concept is derived from the evidence.
>
> 2. Apply his knowledge and intellectual skills so that he can
>   (a) Solve an unfamiliar problem;
>   (b) Analyse relationships;
>   (c) Organize understandings to derive new relationships; and
>   (d) Make valid judgements based on his understanding; and communicate his conclusions in appropriate ways.
>
> (Teacher's Guide, AAS, 1973, pp.9-10)

The ACER, using these content and process objectives as a test grid, has developed batteries of multiple-choice achievement tests for the Web of Life course for grades 11 and 12 in Australian schools. Experienced biology teachers throughout Australia were invited to write items for the tests and to make critical comments on each others test items. The investigator was a member of a review panel of experienced item writers which worked under the direction of the ACER to produce trial forms of each test. Trial testing of the items was carried out in schools taking the Web of Life course and further feedback was obtained from the teachers using these trial materials. Discrimination indices (ϕ coefficient), facility values, and the pattern of responses to the distractors were studied for each item and "after consideration of teacher's opinions and the item analysis data, a selection of the best items was made" for the final forms of the test (ACER, 1972, p.3).

Horn (1966) points out that an achievement test is both a predictor and an assessment, and says that

> in constructing a predictor one must strive to obtain internal consistency among the items, since this tends to ensure that the test is reliable and can correlate with a criterion. In constructing an assessment, on the other hand, one should strive to obtain representativeness of content, whether or not the elements are internally consistent. (p.293)

Ebel (1968), however, points out that test scores need to be reliable in order to be useful for prediction or assessment and that if reliability is considered as the ratio of true score variance to obtained score variance (or operationally, the correlation between measurements of the same characteristic obtained from equivalent but different operations), then internal consistency normally contributes to reliability. He further comments that low internal consistency is often due to faults in the test items such as ambiguity, incorrectly keyed responses or inappropriateness in difficulty. The rejection of such faulty items tends to improve internal consistency and reliability, but Ebel warns that the rejection of items on grounds of item analysis data alone may

> lead to the rejection of items not seriously faulty in themselves but simply different from the other items in what is being measured. The rejection of such items may improve internal consistency but it does nothing to improve reliability, and is almost certain to impair validity. (Ebel, 1968, p.73)

The ACER procedure was to secure high content validity by selecting items for inclusion in the final test forms if they were judged to fairly reflect the processes and content specified, using item analysis data and teachers' opinions about the functioning of the items to revise or reject items which were found to be inadequate.

The two tests used in the present research were Achievement Test AD1 (Diversity: Plants, Animals, and Microorganisms) and Achievement Test AI1 (Inter-relationships) (ACER, 1972).

Sampling of Process Objectives

The two tests AD1 and AI1 were given to three test constructors experienced in the use of the Taxonomy of Educational Objectives (Bloom, et al., 1956) in test construction and analysis. These three judges were asked independently to classify the items in each test according to the major categories of the Taxonomy on the basis of the cognitive skill that appeared to be intended by the test constructor. There was initial disagreement on only three items out of the fifty but after discussion consensus was readily arrived at. The results of the consensus are shown in Table 7.2.1. Bloom's Taxonomy was preferred to the Web of Life classification as a basis for analysing the tests, because the Web of Life classification combines the Bloom categories of Knowledge and Comprehension in asking the student to "state and explain" or "cite, and explain or interpret" a number of aspects of scientific knowledge. "Categories 2 (a) to (d) correspond to 'application', 'analysis', 'synthesis' and 'evaluation' in the

Taxonomy of Educational Objectives ...." (Teacher's Guide, AAS, 1973, p.10).

TABLE 7.2.1

CLASSIFICATION OF ACHIEVEMENT TEST ITEMS BY TAXONOMY[a] CATEGORIES

Taxonomy Category	Test AD1 Number of items	Test AI1 Number of items
1. Knowledge	23	9
2. Comprehension	1	9
3. Application	1	4
4. Analysis	-	3
5. Synthesis	-	-
6. Evaluation	-	-

[a] Bloom, B. S., et al. Taxonomy of Educational Objectives. Handbook 1: Cognitive Domain. New York: Longmans Green, 1956.

It can be seen that test AD1 samples almost exclusively from the Knowledge category, whereas test AI1 samples the higher-order categories of Comprehension, Application, and Analysis.

It was originally intended to use three of the ACER achievement tests but the sections of content completed by all teachers made the administration of the third test inappropriate. Testing usually involves a parsimonious sampling of the intended outcomes and, although the sampling of the cognitive objectives of the Web of Life course would have been better with results from additional tests, the two tests used clearly sample distinctly different elements of course content and of process objectives.

Estimates of Reliability

Estimates of the reliability of the tests are given by the determination of Kuder-Richardson Formula 20 values for two independent samples : one was a sample obtained from the ACER of answer sheets from schools in a number of Australian States, and the other was the sample used in the present research (less thirty-three students from two schools which returned only total test scores for students rather than their responses to all items). The KR 20 values for these samples are given in Table 7.2.2 together with the respective mean scores and standard deviations. Ebel (1968) points out that internal consistency normally gives reasonably good estimates of reliability.

TABLE 7.2.2

ESTIMATES OF RELIABILITY[a] FOR ACHIEVEMENT TESTS

Test	No. of Items	Sample	N	M	SD	KR20
AD1	25	ACER	89	12.55	3.38	0.67
		Present research	113	11.40	3.93	0.79
AI1	25	ACER	113	14.42	3.14	0.58
		Present research	113	14.10	3.21	0.60

[a] Kuder-Richardson Formula 20.

The Kuder-Richardson Formula 20 for estimating test reliability,

$$r = \frac{k}{k-1}\left(1 - \frac{\Sigma\, pq}{\sigma^2}\right)$$

indicates that the value of r depends on k, the number of items in the test, Σpq, the sum of the item variances, and σ^2, the variance of the test scores. The three quantities are interrelated. The greater the number of test items, in general, the larger the values of Σpq and σ^2. Increasing the test length from 25 to 100 items would in itself increase the KR 20 value by 3 percent.

Ebel (1967) has demonstrated that, "while longer tests tend to be more reliable, the quality of the items in them is also influential" (p.125). Table 7.2.3 gives quartile values for the ϕ coefficient of items (calculated from upper and lower thirds) and facility values as reported by the ACER. Although the tests were constructed with a high degree of content validity in mind and are intended to measure a number of skills and a range of content, nevertheless the item discrimination indices are thoroughly acceptable. The facility values, being neither too high nor too low, also reflect an adequate sensitivity of the tests, that is, their ability to spread respondents over the available range of scores (see Gardner, 1975b, pp.14-15).

TABLE 7.2.3

QUARTILE ϕ AND FACILITY VALUES FOR ACHIEVEMENT TEST ITEMS

Test	Sample N[a]	ϕ			Facility		
		Q_1	Mdn	Q_3	Q_1	Mdn	Q_3
AD1	not reported	0.29	0.34	0.47	40	53	64
AI1	not reported	0.23	0.37	0.44	42	67	78

[a] The ACER sample was drawn from schools all over Australia, and although precise data on the sample size is not available it is thought to be 1000. The same sampling procedure was used as for trial testing of previous versions of the tests where samples of 1000 were specified. (ACER, 1972, p.5)

Tests of achievement, such as these, tend to suffer from a lack of variability in scores which, in the case of fully criterion-referenced tests (Popham & Husek, 1969; Glaser & Nitko, 1971), has led numerous authors to propose alternative estimates of their reliability which are not spuriously low (Livingston, 1972; Hambleton & Novik, 1973; Swaminathan, Hambleton, & Algina, 1974, 1975; Huynh, 1976). The values of SD and KR 20 in Table 7.2.2 illustrate Ebel's (1967) comment that "for tests of usual length, it is variation from test to test in the value of σ^2 that is responsible for most of the observed variations in test reliability" (p.125).

Although increasing test length would have had an influence in raising the reliability of the tests it is clear that, to achieve any substantial improvement in reliability, the tests would have had to be more homogeneous or unidimensional with respect to the items. This would have a two fold effect : it would tend to raise the item variances and the variances of the total test scores. However, as argued in chapter 4, on the grounds of the representativeness or ecological validity of the investigation as a whole, it was preferred to use the achievement tests designed for the course, in normal use by classroom teachers, and administered in the teachers' normal schedule of testing.

7.3 MEASURES OF NON-COGNITIVE OUTCOMES

It is recognized by the constructors of the Web of Life course that "public examinations and similar tests are unsuited to testing attitudes and practical skills" (Teacher's Guide, AAS, 1973, p.20). They trust in the nature of the learning experiences to develop these skills and attitudes, without being able to offer help to teachers on how these outcomes are to be measured. Considerable advice is given about drawing up an appropriate test grid and on other facets of test construction, and they observe that written tests "can evaluate content, intellectual processes and the ability to communicate. Other objectives may have to be evaluated in other ways" (Teacher's Guide, AAS, 1976, p.57).

The role of the laboratory

The Teacher's Guide, (AAS, 1973) states that

> the objectives of the course can best be achieved by a course based on laboratory work which involves direct, personal observation and experimentation by the students. It is only by becoming active participants in the processes of science that students can fully appreciate what it is all about. (p.8)

The laboratory work and field investigations are seen to provide learning experiences for the development of inquiry skills; the development of practical skills such as observation, recording data, writing reports, handling apparatus and living materials; learning to work effectively with others; and the development of a number of favourable attitudes.

It has been observed by Grobman (1970) that

> there has been little testing which requires actual performance in a real situation, or in a simulated situation which approaches reality ... to determine not whether a student can verbalize a correct response but whether he can perform an operation, e.g. a laboratory experiment. (pp.192-193)

In the 1940s and 1950s there was an attempt to make the public examinations in Leaving Certificate Biology in Victoria less bookish by including a practical component into the examinations. The Chief Examiner was paid a small sum of money per student to provide some biological material in the examination about which questions would be asked. Demands for testing a wider range of skills with a more comprehensive selection of material brought considerations of expense and logistics which forced abandonment of the program. (Lees, Note 4)

There are a number of examples of laboratory tests in science to be found in the literature on research on science teaching (Kruglak, 1953, 1954, 1955; Frings and Hichar, 1958; Yager, Engen, & Snider, 1959; Coulter, 1966; Pella & Sherman, 1969; Robinson, 1969), but most have been restricted to the measurement of manipulative skills (Shulman & Tamir, 1973). More recently however Tamir and Glassman (1970, 1971a, 1971b) have developed a type of laboratory examination in biology which, it is claimed, "reflects the inquiry objectives of the BSCS philosophy and provides a valid and reliable measure of problem solving ability in a practical laboratory setting" (Tamir, 1974, p.25).

Desirable as it would have been to have had such measures in the present investigation, the administration of such an examination was beyond the financial and logistic resources of the investigator. A future replication of this research may find this possibility appealing if sufficient funds are available.

The measurement of attitudes

As observed in Krathwohl (1964) there is an implicit assumption behind much teaching that if the cognitive objectives are achieved there will be a corresponding development of the appropriate interests and attitudes. Neidt and Hedlund (1969), and Schock (1973) have provided empirical evidence that this assumption is ill-founded. Although affective objectives are admitted to be of fundamental importance to the Web of Life course, as noted, little help is offered to teachers on how progress towards these objectives may be assessed.

In the Teacher's Guide (AAS, 1976) there is an explicit statement of the affective objectives of the Web of Life course:

> It is hoped that, as a result of their experiences gained through the course, students will have maintained or developed:
> (a) An interest in the methods, content and applications of science;
> (b) A lively curiosity, and a willingness to investigate systematically, objectively and carefully, to deliberately suspend judgement, be open to new ideas, be prepared to evaluate ideas in the light of evidence and discard ideas inconsistent with evidence; and
> (c) Some commitment to the value to the individual and community of scientific approaches to problem-solving. (p.21)

In addition there are numerous statements in the course materials in which affective objectives are implicit. In a lengthy section on Teaching Methods for the Biology Teacher in the Teacher's Guide (AAS, 1976) teachers are enjoined to teach biology as inquiry and to avoid a "rhetoric of conclusions" (Schwab, 1962). Through the suggested learning activities, students are to be actively participating in the discovery of the concepts known as 'single propositions' and to view biology learning as an exercise in inquiry itself. In referring to the value of laboratory work or field excursions the Teacher's Guide (AAS, 1973) asserts that

> participation in group work directed towards scientific investigation and solution of problems of mutual concern provide excellent opportunities for learning to work effectively with others, as well as providing special opportunities for developing understanding of the living world and of science and for behaving as scientists themselves behave. (p.8)

There is a striking similarity between the affective objectives expressed in the numerous science curriculum development projects in the USA in the 1960s. This similarity is not altogether surprising since it was a vigourous period of curriculum development accompanied by much public and professional debate. The new curricula were pervasively influenced by similar sets of influences such as the contributions of Bruner (1960) and Schwab (1964); the move to express objectives in behavioural terms (Bloom, 1956; Krathwohl, Bloom, & Masia, 1964; Mager, 1962); and the developmental psychology of Piaget (1947). This similarity of affective aims extended to the Physical Science Study Committee (PSSC) physics course and the BSCS-based Web of Life biology course which were widely adopted in Australia. In 1967 Gardner developed four attitude scales for a curriculum evaluation project in physics (Mackay, 1970, 1971a). The development of these scales is described in detail, Gardner (1972) and given the collective name of the Physics Attitude Index (PAI). The four scales provided measures of students':

(1) views towards physics learning;

(2) beliefs about physics as a process;

(3) views towards scientists; and

(4) enjoyment of physics.

These four dimensions are so consonant with the expressed objectives of the Web of Life course that Lucas adapted them for use in his investigation of changes in content-free skills, knowledge, and attitudes during use of the Web of Life biology course in South Australia (Lucas & Broadhurst, 1972). The forty questions in the PAI were modified by replacing all references to 'physics' by 'biology', and including biological examples or concepts in place of physics examples and concepts. Both instruments were designed for use with students in grades 11 and 12 and care was taken to replace physical examples with biological ones of equal familiarity and similar grammatical structure.

The PAI was developed as four Likert (1932) scales, consisting of a number of opinion statements to which the respondents are asked to indicate whether or not they Strongly Agree, Agree, Not Sure, Disagree, or Strongly Disagree. The responses are scored 5,4,3,2,1, and for unfavourable statements these weightings are reversed. The respondent's score is the sum of the scores he receives on each item. In the PAI the Not Sure response was placed last. Gardner (1972) argues that this "would tend to reduce the frequency of such responses, and increase the standard deviation and hence the reliability of the scores" (p.33). This practice was followed in the biology adaptations of the instrument. Lucas used a slightly different nomenclature to describe the scales and placed four items on scales different to those to which they were allocated in the PAI. A comparison of the allocation of items to the scales in the Lucas' adaptation and the PAI is given in Table 7.3.1.

Gardner (1975a, 1975b) discusses methodological issues of attitude scale construction and the various procedures for allocating items to sub-scales including the use of factor analysis to "provide a further basis for allocating items to scales, for verifying the uniqueness of the various constructs, or for reducing the number of scales when some of them are shown to be redundant" (1975a, p.104). Gardner (1972) examined the PAI for the sensitivity of the scales and their internal consistency, and carried out a minimum residuals factor analysis (Comrey, 1962; Comrey & Ahumada, 1964) with varimax rotation of factors

(Kaiser, 1958) on each scale. With the exception of PAI scale 2, on which students tend to show strongly positive scores, he finds the scores cover most of the available range for scales 1,3, and 4, and concludes that

> there is a substantial body of evidence reflecting the internal consistency of the items within each scale, and unequivocal support for the contention that the four attitude scales in the PAI are measuring unique variables. (p.177)

TABLE 7.3.1

COMPARISONS OF SCALES OF PAI AND LUCAS' ADAPTATION

Instrument	Name of scale	Items
PAI	Views towards physics learning	1,6,7,8,9,11,16,19,28,29
Lucas' scale	Views about methods of teaching biology	1,6,7,8,9,10,11,16,18,19,28,29
PAI	Beliefs about physics as a process	14,20,22,25,27,30,33,36,38,39
Lucas' scale	Views about biology as a process	20,22,25,27,30,33,36,38,39
PAI	Views towards scientists	3,4,13,15,23,31,32,34,35,37
Lucas' scale	Views about scientists	3,4,13,14,15,17,23,31,32,34,35,37
PAI	Enjoyment of physics	2,5,10,12,17,18,21,24,26,40
Lucas' scale	Enjoyment of biology	2,5,12,21,24,26,40

A logical comparison of the items and the description of the scales to which they belong would lead the present investigator to favour the allocation of items as in the PAI rather than Lucas' allocation. A subsequent factor analysis carried out by Lucas caused him to admit that the original allocation to sub-scales as in the PAI was more appropriate than his own (Lucas, Note 5).

In the present research Lucas' adaptation of the items is used and the instrument is called the Biology Attitude Inventory (BAI). The original allocation to sub-scales is adopted. The ten items in each scale follow, preceded by a brief statement defining the affect-outcome being measured.

7.4 THE BIOLOGY ATTITUDE INVENTORY SCALES

Description

BAI scale 1: Views about biology learning (VABL)
(Authoritarian/Non-authoritarian)

The student should view the process of biology learning as a non-authoritarian situation in which students are stimulated to think about biological phenomena, encouraged to 'discover' biological relationships for themselves and to participate in the development of methods for solving problems.

1. If a student was doing an experiment to investigate the effect of bright light on the behaviour of insects, the teacher should not tell him the results; rather the teacher should let the student find out for himself.

6. A biology experiment is best when the instructions are all laid out in the textbook or laboratory manual.

7. The best way for students to learn about bacteria is to listen to a teacher, who knows a great deal about the subject, summarize the facts known about bacteria today.

8. Some laboratory manuals written for biology students to use in their experimental work don't give very definite instructions about what to do; instead they state some problem and expect students to work out their own procedure. This approach is quite good.

9. It is reasonable to ask students to learn laws like Mendel's Laws of Genetics but they should not be told the laws before they do experiments with living organisms, they should try to discover the laws for themselves first.

11. Most experiments should be performed as demonstrations by the teacher, because the teacher is likely to get a better result than the student, and the student would therefore learn more.

16. A biology experiment is good when students are given apparatus and told to solve a problem without instructions of how to do it. That's a real experiment.

19. When you learn a new idea in biology it is very important to thoroughly learn the statement of it so that you know it word perfectly.

28. Some biology teachers occasionally give students problems to solve experimentally, even though the teachers do not know the answer to the problems themselves; this is quite a good idea.

29. The best biology teacher would be the one who tells you as many facts about biology as possible.

BAI scale 2: Views about biology as a process (VABP)
(Closed/Open)

The student should come to view biology as an open rather than a closed process which by its very nature is dynamic, creative, tentative and unfinished.

14. People with imagination would certainly be more successful as writers and artists than as scientists.

20. Once a law in biology is discovered it should not need changing.

22. For every research problem facing the practising biologist there is a well-defined procedure which allows him to obtain the answer.

25. Men may have some of their most cherished ideas shattered by discoveries in biology.

27. Eventually all the laws of biology will be discovered and little further research will be necessary.

30. It is pretty well certain that biologists are reaching the limits of things a man can know.

33. In biology there are many problems for which biologists at present do not even know the procedure for obtaining the answers.

36. It is important for a biologist to be able to throw away widely accepted ideas and think without restriction.

38. There is no end to the new things that a biologist may discover.

39. In the future, people will be surprised at how much biology has changed since now.

BAI scale 3: Views about scientists (VAS) (Eccentric/Normal)

The student should view scientists as normal, active, occasionally fallible human beings who are different from other people only in the area of their special training.

3. Scientists are about as healthy and as physically fit as other people.

4. Scientists are rather eccentric persons who live in a world of their own.

13. Scientists are less warm and friendly than other people.

15. Hardly any scientists believe that God exists.

23. Scientists are interested in religion to the same extent as other people.

31. Scientists are more devoted to their work than most people.

32. Scientists would not be interested in ordinary community activities.

34. Scientists are less interested in art, music and literature than other people.

35. Scientists are peculiar people.

37. The proportion of scientists that are happily married is about the same as in the rest of the population.

BAI scale 4: Enjoyment of biology (EOB) (Antipathy/Commitment)

The student should come to view biology as an important activity for him/herself; he/she should become committed to actively searching for an understanding of biological phenomena, and gain enjoyment thereby.

2. I enjoy reading books about biology.

5. Biology is worth doing whether you want to be a scientist or not.

10. I feel that it is a waste of time and effort for me to investigate any natural phenomenon for which explanations already exist.

12. Biology is just a dull grind for me.

17. I don't see why a biologist would want to spend most of his life doing experiments and analysing results.

18. I like the way biology challenges us to find out why things happen.

21. I cannot see the point in studying topics like producers and

decomposers in any greater detail.

24. I would enjoy having a job in a biology laboratory during one of my school holidays.

26. Learning about the sort of thing that anyone can see happening makes biology boring for me.

40. I find learning about bacteria most enjoyable.

Estimates of reliability

Coefficients of reliability such as the split-half correlation, corrected for length by the Spearman-Brown formula, and the Kuder-Richardson Formula 20 are measures of the internal consistency of the scales, and under normal conditions give good estimates of reliability.

Data on the biology attitude scales are available from Lucas' modification based on a sample from ten grade 12 biology classes in South Australia (the number of students is not given), and for the BAI from the present research on a sample of 146 grade 11 biology students in Victoria. For interest, corrected split-half correlations for the PAI scales are given, these scales containing the equivalent items to the BAI. The PAI data are from Gardner (1972) and are based on a sample of 1014 grade 11 physics students in Victoria. These reliability coefficients are presented in Table 7.4.1

TABLE 7.4.1

ESTIMATES OF RELIABILITY FOR ATTITUDE SCALES

Research	Coefficient	Scale 1 (VABL)	Scale 2 (VABP)	Scale 3 (VAS)	Scale 4 (EOB)
Lucas & Broadhurst	KR20	0.68	0.60	0.61	0.64
Present research	KR20	0.76	0.71	0.83	0.81
Gardner	split-half	0.66	0.60	0.81	0.76
	corrected split-half	0.80	0.75	0.89	0.86

The coefficients for the BAI scales are higher than Lucas and Broadhurst found, which adds additional weight to the claim that the items are more appropriately found in the scales to which they were originally assigned in the physics research (Mackay, 1970, 1971). The values are acceptably high for the purposes of making inferences about groups of students although not sufficiently high for individual prediction.

CHAPTER 8

A MULTIPLE REGRESSION ANALYSIS

8.1 THE INTENT OF THE ANALYSIS

8.2 NON-ORTHOGONALITY OF INDEPENDENT VARIABLES

8.3 NON-ORTHOGONALITY BETWEEN INTERACTIONS AND MAIN EFFECTS

8.4 THE APPROPRIATE ERROR TERM

8.5 THE ATTRIBUTE-TREATMENT INTERACTIONS

8.6 MAIN EFFECTS

8.1 THE INTENT OF THE ANALYSIS

The present study set out to examine the relationships between three classes of student characteristics (personality traits, cognitive preferences and general ability) and contrasting styles of instructional treatment in terms of their outcomes on two measures of achievement and four attitude scales. The 10 personality scales, two cognitive preference scales and the measure of general ability provided continuous data, as did the dependent variables, and the treatment variable was dichotomous.

Students can not be said to have a treatment score: they belong to one treatment group or another. Thus the treatment provides nominal scale data only. Cohen (1968) describes how such data can be treated in multiple regression analysis by using a dummy variable coding procedure. Membership of one category can be shown by assigning an arbitrary number, while non-membership can be shown by assigning another arbitrary number, for example, 1 to members of treatment Group 1 and 0 to members of treatment Group 2. The coding used in this analysis is discussed in section 8.3.

The measures of the other independent variables and the dependent variables are described in detail in the preceding chapters and provide the ordered quantitative data typical of educational research and treated as equal interval scales. Gardner's (1975d) recent review of measurement scales and the use of parametric statistics concludes, inter alia, that such summated scales produce scores that are only mildly distorted versions of interval scales and that parametric procedures are robust and yield valid conclusions even with such mildly distorted data. The data provided for each subject are assumed to represent independent events and the individual is treated as the unit of analysis.

The aim of the analysis was to search for relationships, especially attribute-treatment interactions, among the six dependent variables and 14 independent variables. Each dependent or criterion variable was treated separately and multiple regression procedures were used to analyse the relationships between each dependent variable and the set of independent or predictor variables.

Cohen (1968) and Kerlinger and Pedhazur (1973) have described the virtual identity of multiple regression analysis and analysis of variance. They are both expressions of the general linear model, although most textbooks of statistics in the behavioural sciences treat them as entirely different algorithms. Cohen ascribes this treatment to their different origins and development in the history of behavioural science research.

> [Multiple regression] began to be exploited in the biological and behavioral sciences around the turn of the century in the course of the study of natural variation (Galton, Pearson, Yule). A couple of decades later [fixed-model analysis of variance/covariance] came out of the structure of (agronomic) experimentation, that is, of artificial or experimentally manipulated variation, where the treatments were carefully varied over the experimental material in efficient and logically esthetic experimental designs. (p.426)

The choice of multiple regression analysis in this investigation rather than analysis of variance was based primarily on its capacity to analyze the collective and separate contributions of a large number of independent variables to the variation of a dependent variable. However, there are other advantages in using multiple regression techniques rather than many of the conventional analysis of variance procedures.

Firstly, the continuous variables need not be categorized as in analysis of variance, and so there is no loss of variance in the data and consequent loss of statistical power.

Secondly, multiple regression procedures can handle the combination of continuous variables and the dichotomous variable: the instructional treatment variable can be treated by 'dummy' or 'contrast' coding as described below.

Thirdly, the presentation of the analyses in the form of regression equations indicates clearly the proportions of the total variance in a criterion variable that are being accounted for by each independent variable and so gives an indication of the practical significance of the relationships so described. Such information can be recovered from analysis of variance but is not done so as a matter of course.

Fourthly, multiple regression procedures have advantages in dealing with the non-orthogonality of many naturally-occurring variables. The division of scores into high and low groups (or high, middle and low groups) for analysis of variance in factorial designs results in unequal numbers in cells when variables are correlated. Procedures often used in analysis of variance to equalize cells distort the data by treating the variables as though they were orthogonal when clearly they were not. The procedures adopted in this investigation permit rational and explicit decisions to be made about the redefinition of non-orthogonal variables.

The difficulties associated with multiple regression as a general data-analytic technique relate basically to the unreliability of regression weights, particularly with small samples and large numbers of independent variables, and also to the fact that a different order of entry of independent variables into a regression equation will alter the squared semi-partial correlations.

Where independent variables are orthogonal, where the sample size is large, and where the number of independent variables is small regression weights are more reliable. The measures taken with regard to the non-orthogonality of the independent variables and their reduction in number are described in the next section. There is no unexceptional method for determining the order in which variables should be considered in the regression equation. In some research problems there is no postulated order of importance and, as in stepwise multiple regression procedures, the variable that explains the greatest amount of variance unexplained by the variables already in the equation is entered by the computer at each step, that is, the one with the largest squared partial correlation with the dependent variable. In this investigation, predictors were progressively eliminated in a pre-determined order from a regression equation that contained all predictors. The establishment of the order among the sets of independent variables is also described in the following section.

The initial regression model chosen was:

$$\begin{aligned} Y_j = {} & b_o + b_1 T_j + b_2 I_j + \sum_{i=3}^{12} b_i PS_{(i-2)j} + \sum_{i=13}^{14} b_i CP_{(i-12)j} \\ & + b_{15} T_j.I_j + \sum_{i=16}^{25} b_i T_j.PS_{(i-15)j} + \sum_{i=26}^{27} b_i T_j.CP_{(i-25)j} \\ & + \varepsilon_j \qquad \ldots(1) \end{aligned}$$

where Y is the dependent variable, T is the treatment variable, I is the general ability variable, PS_1, PS_2 ... PS_{10} are the 10 personality scale variables, and CP_1, CP_2 are 2 cognitive preference variables.

It can be seen that the interaction variables were defined by multiplication of the treatment variable with each attribute variable in turn. Quadratic terms were not included in the model because of the limited number of subjects involved. After the reduction in the number of independent variables the ratio of subjects to variables was only 18:1.

The intent of the analysis was to reduce this 'full' model to the simplest possible model which describes the relationships between the data without significant loss of information. The original regression equation contains all attribute and treatment variables and the first order attribute-treatment interactions. These variables were considered in a pre-determined order and those that failed to contribute significantly to the explained variance were eliminated at each successive step.

8.2 NON-ORTHOGONALITY OF INDEPENDENT VARIABLES

The initial regression equation describes scores on the criterion variable in terms of a linear combination of independent components. However, there is reason to believe that non-zero correlations may exist between the attribute variables, that is, the measures of personality, cognitive preference, and general ability. Mackay (1975) found significant relationships between cognitive preferences and both general ability and personality attributes. Heist and Yonge (1962) reported significant correlations between a number of their personality scales which are used in this study and the College Entrance Examination Board Scholastic Aptitude Test. The treatment variable can be assumed to be randomly applied and hence uncorrelated with the three attribute variables.

Non-orthogonality between independent variables is usually treated in analysis of variance by equalizing cells. This results, in effect, in a redefinition of the variables so that they are not now correlated. If this redefinition is not recognized, problems of interpretation can follow (see Humphreys & Dachler's (1969) criticism of Jensen's (1968) use of correlated individual difference variables in a pseudo-orthogonal design). Thus it appears to be more desirable to make explicit those decisions which are taken in the redefinition of the independent variables to achieve orthogonality. For convenience in explanation the personality scales and cognitive preferences are considered initially as single variables.

The non-orthogonality of the three attribute variables I, PS, and CP are represented in Figure 8.2.1.

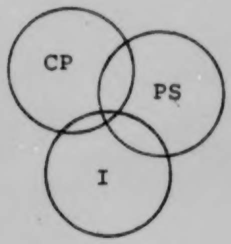


Figure 8.2.1. The non-orthogonality of general ability, cognitive preference, and personality attributes.

Standard texts on psychological testing (e.g. Cronbach, 1970; Tyler, 1975) observe that, judged on their predictive validity, intelligence tests are basically measures of scholastic aptitude or scholastic ability. The known relationship between intelligence and the types of achievement measures used as dependent variables in this study suggests that the first step in the redefinition of the variables should be to separate out the total contribution of general ability as shown in Figure 8.2.2.

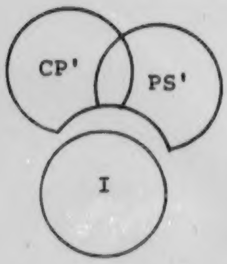


Figure 8.2.2. The separation of the total contribution of general ability from other attribute variables.

I represents the total contribution of the general ability variable, whereas CP' and PS' are those parts of CP and PS that are independent of I.

Because of its expected importance in prediction then, I was considered the most proximal (Snow, 1974) of the variables considered and is the last to be dropped from the regression equation. The variance accounted for by general ability will include not only that due to its unique influence on the dependent variable, but also those proportions due to the overlap with cognitive preferences and personality scales. Overall and Spiegel (1969) discuss three methods of partitioning the sums of squares for the analysis of variance. Their Method 3, which is called the step-down analysis "involves an initial ordering of the effects and then estimating each effect adjusted for those preceding it in the ordering and ignoring those following it" (p.316). The treatment of the main effect of general ability in this investigation follows this strategy. However, the complete strategy for the multiple regression analysis does not follow Method 1, Method 2, or Method 3 (Overall & Spiegel, 1969) but uses features of the various methods (and of Rawlings Method 1', Rawlings, 1973) to resolve specific issues as they arise. Overall, Spiegel, and Cohen (1975) assert that they "should not presume to deny the informed investigator the great flexibility of general linear regression methods for the analysis of experimental data using a variety of models" (p.186). In this analysis care has been taken to be explicit about the models tested, and the bases for the decisions to test these models.

There were no obvious theoretical grounds for postulating an order between cognitive preference and personality scale variables, and the strategy of Overall and Spiegel's Method 1 appeared to be the more appropriate strategy of partitioning the sums of squares between these variables. It was not expected that there would be any substantial correlation between cognitive preferences and the personality scales in the population of subjects from which this sample was drawn. Mackay's (1975) results show only one correlation significant at the $\alpha = 0.05$ level between four cognitive preference scales and two personality dimensions, although the magnitude of this correlation is not given. It was decided, therefore, to ignore the overlap between the two thus giving rise to the three orthogonal variables as shown in Figure 8.2.3.

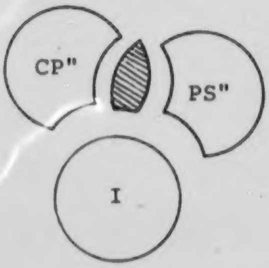


Figure 8.2.3. Redefinition of orthogonal attribute variables.

I represents the total contribution of the general ability variable. CP" represents that part of CP that is independent of I and PS, that is the unique contribution of CP. PS" represents that part of PS that is independent of I and CP, that is, the unique contribution of PS.

There is some loss of statistical power in ignoring the overlap between CP' and PS' if this overlap is related to the dependent variable, but it was expected that this overlap would be small. Subsequent analysis confirmed this expectation. There was only one significant correlation between the four personality factors and two cognitive preference scores ultimately used in this study: an $r = 0.184$ (for $N = 146$, an r of 0.167 is required for significance at the $\alpha = 0.05$ level).

To summarize, the redefined orthogonal attribute variables are:

PS", personality scales controlled for cognitive preferences and general ability,

CP", cognitive preferences controlled for personality scales and general ability, and

I , general ability in its original form.

The testing of the redefined variables I, PS", and CP" can be achieved without the variables being transformed by comparison of the appropriate regression models. The models required are:

(a) $Y = f(T, I, CP, PS)$,

(b) $Y = f(T, I, PS)$,

(c) $Y = f(T, I, CP)$,

(d) $Y = f(T, I)$, and

(e) $Y = f(T)$.

Comparison of models (a) and (b) gives the influence of CP" on Y, comparison of models (a) and (c) gives the influence of PS" on Y, and comparison of models (d) and (e) gives the influence of I on Y.

The comparison of the total sums of squares of the various models permits the testing of the significance of loss in predictive accuracy accompanied by the dropping of terms from the full regression equation

$$F = \frac{(R_1^{\,2} - R_2^{\,2})\ df_n}{R^2 \text{ due to error}/df_d} \qquad \ldots(2)$$

where $R_1^{\,2}$ and $R_2^{\,2}$ are the R^2 associated with the two models being compared.

The definition of the appropriate terms for R^2 due to error, df_n, and df_d will be discussed in a later section.

The interaction terms will be considered in a similar way except for the (I x T) interaction. Unlike I, there is no reason to expect that the general ability x treatment interaction is generally related to cognitive or affective outcomes. Hence only the contribution of the 'unique' part of this interaction variable will be tested. Thus the redefined orthogonal interaction variables are:

(PS" x T), PS x T controlled for (CP" x T), (I" x T) and all main effects,

(CP" x T), CP x T controlled for (PS" x T), (I" x T) and all main effects, and

(I" x T), I x T controlled for (CP" x T), (PS" x T) and all main effects.

These relationships are shown diagrammatically in Figure 8.2.4.

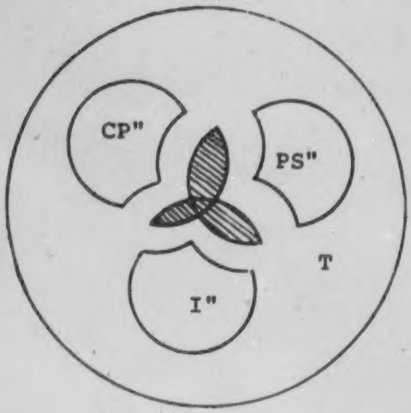


Figure 8.2.4. The orthogonal attribute-treatment interaction variables.

Controlling for main effects to orthogonalize the interactions and main effects provides for a simpler explanation. PS is a simpler variable than PS x T and the interaction term is only interesting if it 'explains' variance in the dependent variable that is not 'explained' by PS and T alone.

To this point in the discussion, PS and CP have been considered as single variables. The CP set consists of two approximately orthogonal cognitive preference variables, reduced from four scales by an unfolding analysis (Coombs, 1958, 1964). Mackay (1972) states that Q-technique factor analyses conducted for sub-samples of the 9263 individuals in his study give an identical solution. Because of the possibility of slight overlap, the two cognitive preference scales were redefined as that part of the particular scale that was independent of the other. As described in chapter 5, personality attributes were measured in this study by 10 OPI scales. Four scales with a substantial number of items overlapping with other scales were omitted. There were substantial correlations between the 10 personality scales, and in order to resolve this non-orthogonality a factor analysis was carried out. The consequent reduction in the number of variables is also favoured by Kerlinger and Pedhazur (1973) in their attack on the use of multiple regression as a shotgun technique. They advise that "before using many variables in multiple regression analysis, some attempt should be made to reduce the number through theory and factor analysis" (p.442). Four orthogonal factors were extracted by using Veldman's (1967) principal components analysis followed by varimax rotation and an eigenvalue cut off of 1.0. The first four roots extracted 72.6% of the trace. Although factor analysis solves the problem of non-orthogonality it often adds to difficulties in interpretation. Fortunately in this study the interpretation of the four factors was not unduly difficult. The factor matrix is relatively simple with 6 of the 10 scales having only one significant non-zero loading.

The redefinition of the attribute variables and the attribute-treatment interaction terms produces the following set of redefined orthogonal variables that were tested in the analysis:

$PF"_1$, scores on personality factor 1 controlled for the other three personality factors ($PF"_2$, $PF"_3$, $PF"_4$), the two cognitive preference scales ($CP"_1$, $CP"_2$), and general ability (I),

$PF"_2$, scores on personality factor 2 controlled for the other three personality factors ($PF"_1$, $PF"_3$, $PF"_4$), the two cognitive preference scales ($CP"_1$, $CP"_2$), and general ability (I),

$PF"_3$, scores on personality factor 3 controlled for the other three personality factors ($PF"_1$, $PF"_2$, $PF"_4$), the two cognitive preference scales ($CP"_1$, $CP"_2$),

	and general ability (I),
PF''_4,	scores on personality factor 4 controlled for the other three personality factors (PF''_1, PF''_2, PF''_3), the two cognitive preference scales (CP''_1, CP''_2) and general ability (I),
CP''_1,	scores on cognitive preference scale 1 controlled for CP''_2, the four personality factors (PF''_1, PF''_2, PF''_3, PF''_4), and general ability (I),
CP''_2,	scores on cognitive preference scale 2 controlled for CP''_1, the four personality factors (PF''_1, PF''_2, PF''_3, PF''_4), and general ability (I),
I,	the raw scores on the ACER B 40 test,
(PF''_1 x T),	scores on personality factor 1 x treatment score controlled for (PF''_2 x T), (PF''_3 x T), (PF''_4 x T), (CP''_1 x T), (CP''_2 x T), (I" x T), and all main effects,
(PF''_2 x T),	scores on personality factor 2 x treatment score controlled for (PF''_1 x T), (PF''_3 x T), (PF''_4 x T), (CP''_1 x T), (CP''_2 x T), (I" x T), and all main effects,
(PF''_3 x T),	scores on personality factor 3 x treatment score controlled for (PF''_1 x T), (PF''_2 x T), (PF''_4 x T), (CP''_1 x T), (CP''_2 x T), (I" x T), and all main effects,
(PF''_4 x T),	scores on personality factor 4 x treatment score controlled for (PF''_1 x T), (PF''_2 x T), (PF''_3 x T), (CP''_1 x T), (CP''_2 x T), (I" x T), and all main effects,
(CP''_1 x T),	scores on cognitive preference scale 1 x treatment score controlled for (PF''_1 x T), (PF''_2 x T),

$(PF''_3 \times T)$, $(PF''_4 \times T)$, $(CP''_2 \times T)$, $(I'' \times T)$, and all main effects,

$(CP''_2 \times T)$, scores on cognitive preference scale 2 x treatment score controlled for $(PF''_1 \times T)$, $(PF''_2 \times T)$, $(PF''_3 \times T)$, $(PF''_4 \times T)$, $(CP''_1 \times T)$, $(I'' \times T)$, and all main effects, and

$(I'' \times T)$, general ability score x treatment score controlled for $(PF''_1 \times T)$, $(PF''_2 \times T)$, $(PF''_3 \times T)$, $(PF''_4 \times T)$, $(CP''_1 \times T)$, $(CP''_2 \times T)$, and all main effects.

As a result of this redefinition of variables the full regression model now becomes:

$$Y_j = b_0 + b_1 T_j + b_2 I_j + \sum_{i=3}^{6} b_i PF''_{(i-2)j} + \sum_{i=7}^{8} b_i CP''_{(i-6)j} + b_9 T_j.I''_j + \sum_{i=10}^{13} b_i T_j.PF''_{(i-9)j} + \sum_{i=14}^{15} b_i T_j.CP''_{(i-13)j} + \varepsilon_j \qquad ...(3)$$

8.3 NON-ORTHOGONALITY BETWEEN INTERACTIONS AND MAIN EFFECTS

The strategy used to test interaction terms so that they are orthogonal to their related main effects has already been considered. However, an essentially similar problem exists in testing for main effects when significant interactions have been found. For example, if the effect of $(PF''_1 \times T)$ was found to be significant, then it must remain in the full model when testing for main effects. When the PF''_1 main effect is tested, its overlap with $(PF''_1 \times T)$ will remain in the full model so that only PF''_1 controlled for $(PF''_1 \times T)$ will be tested. This problem can be overcome if the definition of T is such that ($PF''_1 \times T$) is actually independent of PF''_1. This independence is achieved in the computation of orthogonal contrasts in analysis of variance by assigning the value 1 to subjects in one treatment group and the value -1 to subjects in the other: a procedure described by Cohen (1968) as contrast coding. The simulation of the equal cells analysis of variance by multiple regression analysis requires orthogonality between all terms of the regression equation. Where cells are equal, there is essential agreement by Overall and Spiegel (1969, 1973), Gocka (1973) and Overall, Spiegel, and Cohen (1975) that the nominal scale variables should be coded (1, 0,..., -1), which in the case of contrasting treatments is (1, -1). When this coding is used, the product variable ($PF'' \times T$) will be independent of PF". When the cell numbers are not equal the (1, 0,..., -1) coding does not produce orthogonality (Gocka, 1973; McArthur, Fraser, & West, 1975). In this investigation the difference in cell size is not large (79, 67) and although the (1, -1) coding will leave a slight correlation between the terms, it is believed that the consequent loss of power will be slight.

The consequent definition of the treatment variable was $T = 1$ for groups 2, 3, 8, and $T = -1$ for groups 1, 3, 4, 5, 6. It is clear, after the event, that the groups of variables have been virtually orthogonal and Overall and Spiegel's (1969) Method 1 effect coding could have been adopted in the analysis.

8.4 THE APPROPRIATE ERROR TERM

Orthogonality between terms in a regression model requires not only an appropriate method of coding nominal scale predictor variables, but also the choice of an appropriate algorithm for the computation of the appropriate F statistic.

The denominator in equation 2 contains the R^2 due to error. The usual procedure in analysis of variance is to include all effects, whether significant or non-significant, in the computation of the error term. In multiple regression terms this means using $1 - R^2_{full}$ where R^2_{full} is associated with the model containing all terms (Overall & Spiegel, 1969, p.316; Kim & Kohout, 1975, p.337). When a large number of terms is involved, even if they are not significant, they can act to substantially reduce the error term and so inflate the value of F.

In this analysis the full regression model was used to compute the error term for the test of interaction effects. However, at the completion of that analysis the strategy suggested by Rawlings' (1973) Method 1' was followed: the full model was revised so that it included only those interaction terms that were significant. This revised model was used to compute the term for the main effects analysis.

Thus the algorithm for the interaction effects was

$$F = \frac{(R^2_A - R^2_B)/(V_A - V_B)}{(1 - R^2_{full})/(N - 1 - V_{full})} \qquad \ldots(4)$$

where, models A and B are the models being compared, the full model is the model described by equation 3, V_A is the number of variables in model A, V_B is the number of variables in model B, and V_{full} is the number of variables in the full model.

The algorithm used for testing main effects was

$$F = \frac{(R^2_A - R^2_B)/(V_A - V_B)}{(1 - R^2_C)/(N - 1 - V_C)} \qquad \ldots(5)$$

where, model C contains all main effects and only those interaction terms found to be significant, and V_C is the number of variables in model C.

8.5 THE ATTRIBUTE-TREATMENT INTERACTIONS

The multiple regression analysis was carried out using the subprogram 'Regression' (Kim & Kohout, 1975) from the Statistical Package for the Social Sciences, SPSS (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975).

The models used for comparison, the appropriate squared semi-partial correlations, and the F ratios determined are shown in Tables 8.5.1, 8.5.2, and 8.5.3. The algorithm used for the computation of the F ratios was equation 4.

The models are described as follows:

a.1, full model (equation 3),

a.2, full model with the ($PF''_1 \times T$) term removed,

a.3, full model with the ($PF''_2 \times T$) term removed,

a.4, full model with the ($PF''_3 \times T$) term removed,

a.5, full model with the ($PF''_4 \times T$) term removed,

a.6, full model with the ($CP''_1 \times T$) term removed,

a.7, full model with the ($CP''_2 \times T$) term removed, and

a.8, full model with the ($I'' \times T$) term removed,

where 'a' represents the criterion variable being considered. In each case R^2_{full} is model A, $(V_A - V_B) = 1$, and $V_{full} = 15$.

The six dependent variables, described in Chapter 7, are:

1, BAI scale 1, Views about biology learning (VABL),

2, BAI scale 2, Views about biology as a process (VABP),

3, BAI scale 3, Views about scientists (VAS),

4, BAI scale 4, Enjoyment of biology (EOB),

5, Achievement test AD1, Diversity (of living organisms), and

6, Achievement test AI1, Inter-relationships (among living organisms).

TABLE 8.5.1

VALUES OF R^2 AND F FOR (PERSONALITY FACTOR x TREATMENT) INTERACTIONS

Dependent Variable	Independent Variable	Model A	Model B	R^2_A	R^2_B	$R^2_A - R^2_B$	F
VABL	(PF"$_1$ x T)	1.1	1.2	.1988	.1858	.0130	2.11
VABL	(PF"$_2$ x T)	1.1	1.3	.1988	.1988	.0000	0.00
VABL	(PF"$_3$ x T)	1.1	1.4	.1988	.1952	.0036	0.58
VABL	(PF"$_4$ x T)	1.1	1.5	.1988	.1876	.0112	1.82
VABP	(PF"$_1$ x T)	2.1	2.2	.2378	.2344	.0034	0.58
VABP	(PF"$_2$ x T)	2.1	2.3	.2378	.2375	.0003	0.05
VABP	(PF"$_3$ x T)	2.1	2.4	.2378	.2345	.0033	0.59
VABP	(PF"$_4$ x T)	2.1	2.5	.2378	.2098	.0280	4.78*
VAS	(PF"$_1$ x T)	3.1	3.2	.1233	.1187	.0046	0.68
VAS	(PF"$_2$ x T)	3.1	3.3	.1233	.1001	.0232	3.44
VAS	(PF"$_3$ x T)	3.1	3.4	.1233	.1231	.0002	0.03
VAS	(PF"$_4$ x T)	3.1	3.5	.1233	.1233	.0000	0.00
EOB	(PF"$_1$ x T)	4.1	4.2	.1887	.1887	.0000	0.00
EOB	(PF"$_2$ x T)	4.1	4.3	.1887	.1874	.0013	0.21
EOB	(PF"$_3$ x T)	4.1	4.4	.1887	.1875	.0012	0.19
EOB	(PF"$_4$ x T)	4.1	4.5	.1887	.1810	.0077	1.23
AD1	(PF"$_1$ x T)	5.1	5.2	.3719	.3718	.0001	0.02
AD1	(PF"$_2$ x T)	5.1	5.3	.3719	.3706	.0013	0.27
AD1	(PF"$_3$ x T)	5.1	5.4	.3719	.3705	.0014	0.29
AD1	(PF"$_4$ x T)	5.1	5.5	.3719	.3459	.0260	5.38*
AI1	(PF"$_1$ x T)	6.1	6.2	.3275	.3274	.0001	0.02
AI1	(PF"$_2$ x T)	6.1	6.3	.3275	.3187	.0088	1.70
AI1	(PF"$_3$ x T)	6.1	6.4	.3275	.3062	.0213	4.12*
AI1	(PF"$_4$ x T)	6.1	6.5	.3275	.3239	.0036	0.76

* $p < .05$

Critical $F_{1,130}$ = 3.92 at 0.05 level of confidence

TABLE 8.5.2

VALUES OF R^2 AND F FOR (COGNITIVE PREFERENCE x TREATMENT) INTERACTIONS

Dependent Variable	Independent Variable	Model A	Model B	R^2_A	R^2_B	$R^2_A - R^2_B$	F
VABL	(CP"$_1$ x T)	1.1	1.6	.1988	.1897	.0091	1.48
VABL	(CP"$_2$ x T)	1.1	1.7	.1988	.1884	.0104	1.69
VABP	(CP"$_1$ x T)	2.1	2.6	.2378	.2078	.0300	5.12*
VABP	(CP"$_2$ x T)	2.1	2.7	.2378	.2371	.0007	0.12
VAS	(CP"$_1$ x T)	3.1	3.6	.1233	.1210	.0023	0.34
VAS	(CP"$_2$ x T)	3.1	3.7	.1233	.1129	.0104	1.54
EOB	(CP"$_1$ x T)	4.1	4.6	.1887	.1647	.0240	3.85
EOB	(CP"$_2$ x T)	4.1	4.7	.1887	.1884	.0003	0.05
AD1	(CP"$_1$ x T)	5.1	5.6	.3719	.3608	.0111	2.30
AD1	(CP"$_2$ x T)	5.1	5.7	.3719	.3322	.0397	8.22**
AI1	(CP"$_1$ x T)	6.1	6.6	.3275	.3268	.0007	0.16
AI1	(CP"$_2$ x T)	6.1	6.7	.3275	.3147	.0128	2.47

* $p < .05$; ** $p < .01$

Critical $F_{1,130}$ = 3.92 at 0.05 level of confidence

TABLE 8.5.3

VALUES OF R^2 AND F FOR (GENERAL ABILITY x TREATMENT) INTERACTIONS

Dependent Variable	Independent Variable	Model A	Model B	R^2_A	R^2_B	$R^2_A - R^2_B$	F
VABL	(I" x T)	1.1	1.8	.1988	.1906	.0082	1.33
VABP	(I" x T)	2.1	2.8	.2378	.2371	.0007	0.12
VAS	(I" x T)	3.1	3.8	.1233	.1213	.0020	0.30
EOB	(I" x T)	4.1	4.8	.1887	.1887	.0000	0.00
AD1	(I" x T)	5.1	5.8	.3719	.3716	.0003	0.06
AI1	(I" x T)	6.1	6.8	.3275	.3104	.1071	3.31

Critical $F_{1,130}$ = 3.92 at 0.05 level of confidence

8.6 THE MAIN EFFECTS

As a result of the foregoing analysis of the interactions, the full model for each criterion variable was simplified by dropping out those terms that did not significantly contribute to the explanatory power of the regression equation (Rawlings, 1973). The simplified models for the criterion variables are respectively:

Y_1 = f(main effects),

Y_2 = f(main effects, $PF''_4 \times T$, $CP''_1 \times T$),

Y_3 = f(main effects),

Y_4 = f(main effects),

Y_5 = f(main effects, $PF''_4 \times T$, $CP''_2 \times T$), and

Y_5 = f(main effects, $PF''_3 \times T$).

The simplified models for each dependent variable were now used as the full model for the analysis of main effects as described in section 8.4. The test of effects was described in section 8.2. Thus the appropriate models are:

a.9, the simplified model for each dependent variable as just described,

a.10, the simplified model with the CP''_1 term removed,

a.11, the simplified model with the CP''_2 term removed,

a.12, the simplified model with the PF''_1 term removed,

a.13, the simplified model with the PF''_2 term removed,

a.14, the simplified model with the PF''_3 term removed,

a.15, the simplified model with the PF''_4 term removed,

a.16, the simplified model with the treatment term removed,

a.17, the model containing significant interactions, T and I only, and

a.18, the model containing significant interactions and T only, where 'a' represents the criterion variable being considered.

The models used for comparison, the appropriate squared semipartial correlations, the appropriate degrees of freedom, and the F ratios determined are shown in Tables 8.6.1 to 8.6.6. The algorithm used for the computation of the F ratios was equation 5.

Taking into account the significant main effects as well as the significant interaction effects, the final models for each criterion variable are:

$$Y_1 = f(PF''_3, PF''_4, I)$$

$$Y_2 = f(T, PF''_3, I, PF''_4 \times T, CP''_1 \times T)$$

$$Y_3 = f(T)$$

$$Y_4 = f(PF''_1, PF''_3)$$

$$Y_5 = f(T, PF''_3, PF''_4, I, PF''_4 \times T, CP''_2 \times T)$$

$$Y_6 = f(T, PF''_1, PF''_4, I, PF''_3 \times T).$$

For the reader's convenience, the nomenclature PF''_1, and so on, will be presented as PF1 in the ensuing chapters.

TABLE 8.6.1

VALUES OF R^2 AND F FOR MAIN EFFECTS: DEPENDENT VARIABLE 1 (VABL)

Dependent Variable	Independent Variable	Model A	R^2_A	V_A	Model B	R^2_B	V_B	Model C	R^2_C	V_C	F
VABL	CP''_1	1.9	.1540	8	1.10	.1540	7	1.9	.1540	8	0.00
VABL	CP''_2	1.9	.1540	8	1.11	.1502	7	1.9	.1540	8	0.62
VABL	PF''_1	1.9	.1540	8	1.12	.1540	7	1.9	.1540	8	0.00
VABL	PF''_2	1.9	.1540	8	1.13	.1461	7	1.9	.1540	8	1.30
VABL	PF''_3	1.9	.1540	8	1.14	.1162	7	1.9	.1540	8	6.12*
VABL	PF''_4	1.9	.1540	8	1.15	.1028	7	1.9	.1540	8	8.29**
VABL	Treatment	1.9	.1540	8	1.16	.1528	7	1.9	.1540	8	0.19
VABL	IQ	1.17	.0522	2	1.18	.0034	1	1.9	.1540	8	7.90**

* $p < .05$; ** $p < .01$

Critical $F_{1,137} = 3.92$ at .05 level of confidence

TABLE 8.6.2

VALUES OF R^2 AND F FOR MAIN EFFECTS: DEPENDENT VARIABLE 2 (VABP)

Dependent Variable	Independent Variable	Model A	R^2_A	V_A	Model B	R^2_B	V_B	Model C	R^2_C	V_C	F
VABP	CP''_1	2.9	.2305	10	2.10	.2216	9	2.9	.2305	10	1.56
VABP	CP''_2	2.9	.2305	10	2.11	.2305	9	2.9	.2305	10	0.00
VABP	PF''_1	2.9	.2305	10	2.12	.2305	9	2.9	.2305	10	0.00
VABP	PF''_2	2.9	.2305	10	2.13	.2304	9	2.9	.2305	10	0.02
VABP	PF''_3	2.9	.2305	10	2.14	.1048	9	2.9	.2305	10	22.05***
VABP	PF''_4	2.9	.2305	10	2.15	.2304	9	2.9	.2305	10	0.02
VABP	Treatment	2.9	.2305	10	2.16	.1992	9	2.9	.2305	10	5.49*
VABP	IQ	2.17	.1231	2	2.18	.0743	1	2.9	.2305	10	8.56**

* $p < .05$; ** $p < .01$; *** $p < .001$

Critical $F_{1.135} = 3.92$ at .05 level of confidence

TABLE 8.6.3

VALUES OF R^2 AND F FOR MAIN EFFECTS: DEPENDENT VARIABLE 3 (VAS)

Dependent Variable	Independent Variable	Model A	R^2_A	V_A	Model B	R^2_B	V_B	Model C	R^2_C	V_C	F
VAS	CP"$_1$	3.9	.08203	8	3.10	.07965	7	3.9	.08203	8	0.36
VAS	CP"$_2$	3.9	.08203	8	3.11	.06591	7	3.9	.08203	8	2.41
VAS	PF"$_1$	3.9	.08203	8	3.12	.07961	7	3.9	.08203	8	0.36
VAS	PF"$_2$	3.9	.08203	8	3.13	.08163	7	3.9	.08203	8	0.06
VAS	PF"$_3$	3.9	.08203	8	3.14	.06059	7	3.9	.08203	8	3.20
VAS	PF"$_4$	3.9	.08203	8	3.15	.07619	7	3.9	.08203	8	0.87
VAS	Treatment	3.9	.08203	8	3.16	.05178	7	3.9	.08203	8	4.51*
VAS	IQ	3.17	.03837	2	3.18	.03311	1	3.9	.08203	8	0.79

* $p < .05$

Critical $F_{1,137}$ = 3.92 at .05 level of confidence

TABLE 8.6.4

VALUES OF R^2 AND F FOR MAIN EFFECTS: DEPENDENT VARIABLE 4 (EOB)

Dependent Variable	Independent Variable	Model A	R^2_A	V_A	Model B	R^2_B	V_B	Model C	R^2_C	V_C	F
EOB	CP''_1	4.9	.1555	8	4.10	.1466	7	4.9	.1555	8	1.44
EOB	CP''_2	4.9	.1555	8	4.11	.1524	7	4.9	.1555	8	0.50
EOB	PF''_1	4.9	.1555	8	4.12	.1216	7	4.9	.1555	8	5.50*
EOB	PF''_2	4.9	.1555	8	4.13	.1402	7	4.9	.1555	8	2.48
EOB	PF''_3	4.9	.1555	8	4.14	.0782	7	4.9	.1555	8	12.50***
EOB	PF''_4	4.9	.1555	8	4.15	.1515	7	4.9	.1555	8	0.65
EOB	Treatment	4.9	.1555	8	4.16	.1524	7	4.9	.1555	8	0.50
EOB	IQ	4.17	.00927	2	4.18	.00413	1	4.9	.1555	8	0.83

* $p < .05$; ** $p < .01$; *** $p < .001$

Critical $F_{1,137} = 3.92$ at .05 level of confidence

TABLE 8.6.5

VALUES OF R^2 AND F FOR MAIN EFFECTS: DEPENDENT VARIABLE 5 (AD1)

Dependent Variable	Independent Variable	Model A	R^2_A	V_A	Model B	R^2_B	V_B	Model C	R^2_C	V_C	F
AD1	CP''_1	5.9	.3589	10	5.10	.3587	9	5.9	.3589	10	0.04
AD1	CP''_2	5.9	.3589	10	5.11	.3573	9	5.9	.3589	10	0.34
AD1	PF''_1	5.9	.3589	10	5.12	.3518	9	5.9	.3589	10	1.50
AD1	PF''_2	5.9	.3589	10	5.13	.3580	9	5.9	.3589	10	0.19
AD1	PF''_3	5.9	.3589	10	5.14	.3383	9	5.9	.3584	10	4.34*
AD1	PF''_4	5.9	.3589	10	5.15	.3150	9	5.9	.3589	10	9.24**
AD1	Treatment	5.9	.3589	10	5.16	.3318	9	5.9	.3589	10	5.71*
AD1	IQ	5.17	.2858	2	5.18	.1656	1	5.9	.3589	10	25.31***

* $p < .05$;   ** $p < .01$;   *** $p < .001$

Critical $F_{1,135}$ = 3.92 at .05 level of confidence

TABLE 8.6.6

VALUES OF R^2 AND F FOR MAIN EFFECTS: DEPENDENT VARIABLE 6 (AI1)

Dependent Variable	Independent Variable	Model A	R^2_A	V_A	Model B	R^2_B	V_B	Model C	R^2_C	V_C	F
AI1	CP''_1	6.9	.2844	9	6.10	.2843	8	6.9	.2844	9	0.02
AI1	CP''_2	6.9	.2844	9	6.11	.2837	8	6.9	.2844	9	0.13
AI1	PF''_1	6.9	.2844	9	6.12	.2629	8	6.9	.2844	9	4.09*
AI1	PF''_2	6.9	.2844	9	6.13	.2803	8	6.9	.2844	9	0.78
AI1	PF''_3	6.9	.2844	9	6.14	.2811	8	6.9	.2844	9	0.63
AI1	PF''_4	6.9	.2844	9	6.15	.2560	8	6.9	.2844	9	5.40*
AI1	Treatment	6.9	.2844	9	6.16	.2295	8	6.9	.2844	9	10.43**
AI1	IQ	6.17	.2279	2	6.18	.1131	1	6.9	.2844	9	21.66***

* $p < .05$; ** $p < .01$; *** $p < .001$

Critical $F_{1,136}$ = 3.92 at .05 level of confidence

CHAPTER 9

INTERPRETATION OF FINDINGS

9.1 THE FINAL REGRESSION MODELS

9.2 STATISTICAL INFERENCE

9.3 THE FINAL TREND EQUATIONS

9.4 VIEWS ABOUT BIOLOGY LEARNING (VABL)

9.5 VIEWS ABOUT BIOLOGY AS A PROCESS (VABP)

9.6 VIEWS ABOUT SCIENTISTS (VAS)

9.7 ENJOYMENT OF BIOLOGY (EOB)

9.8 ACHIEVEMENT TEST AD1 (DIVERSITY OF LIVING ORGANISMS)

9.9 ACHIEVEMENT TEST AI1 (INTER-RELATIONSHIPS)

9.1 THE FINAL REGRESSION MODELS

For each criterion variable the full regression model has been simplified by dropping out, firstly, those interaction terms and, secondly, those main effects terms that did not significantly contribute to the explanatory power of the regression equation.

Table 9.1.1 shows the F values for each criterion variable using treatment, the seven attribute variables, and seven attribute-treatment interactions as individual predictors, controlling for all others.

It can be seen that the most parsimonious statement of relationships between each dependent variable and the independent variables considered in the full regression model is as follows:

1. $Y'_{VABLj} = b_0 + b_1\, PF3''_j + b_2\, PF4''_j + b_3\, I_j + \varepsilon_j$

2. $Y'_{VABPj} = b_0 + b_1\, T_j + b_2\, PF3''_j + b_3\, I_j + b_4\, T_j.PF4''_j + b_5\, T_j.CP1''_j + \varepsilon_j$

3. $Y'_{VASj} = b_0 + b_1\, T_j + \varepsilon_j$

4. $Y'_{EOBj} = b_0 + b_1\, PF1''_j + b_2\, PF3''_j + \varepsilon_j$

5. $Y'_{AD1j} = b_0 + b_1\, T_j + b_2\, PF3''_j + b_3\, PF4''_j + b_4\, I_j + b_5\, T_j.PF4''_j + b_6\, T_j.CP2''_j + \varepsilon_j$

6. $Y'_{AI1j} = b_0 + b_1\, T_j + b_2\, PF1''_j + b_3\, PF4''_j + b_4\, I_j + b_5\, T_j.PF3''_j + \varepsilon_j$

It can be seen from Table 9.1.1 that attribute-treatment interactions, as might be expected from the literature review, are not common, comprising 5 of the 42 interaction effects tested. However, this is substantially more than expected as chance occurrence using the $\alpha = 0.05$ level for tests of significance.

Inspection of Table 9.1.2 also shows that for those criterion variables where interactions have been found, the final prediction equation accounts for substantially more of the variance than for other

TABLE 9.1.1

F VALUE FOR EACH CRITERION VARIABLE USING TREATMENT, SEVEN ATTRIBUTE VARIABLES AND SEVEN ATTRIBUTE-TREATMENT INTERACTIONS AS INDIVIDUAL PREDICTORS, CONTROLLING FOR ALL OTHERS.

Criterion	T X PF1	T x PF2	T x PF3	T x PF4	T x CP1	T x CP2	T x I	PF1	PF2	PF3	PF4	CP1	CP2	T	I
VABL	2.11	0.00	0.58	1.82	1.48	1.69	1.33	0.00	1.30	6.12*	8.29**	0.00	0.62	0.19	7.90**
VABP	0.58	0.05	0.59	4.78*	5.12*	0.12	0.12	0.00	0.02	22.05***	0.02	1.56	0.00	5.49*	8.56**
VAS	0.68	3.44	0.03	0.00	0.34	1.54	0.30	0.36	0.06	3.20	0.87	0.36	2.41	4.51*	0.79
EOB	0.00	0.21	0.19	1.23	3.85	0.05	0.00	5.50*	2.48	12.50***	0.65	1.44	0.50	0.50	0.83
AD1	0.02	0.27	0.29	5.38*	2.30	8.22**	0.06	1.50	0.19	4.34*	9.24**	0.04	0.34	5.71*	25.31***
AI1	0.02	1.70	4.12*	0.76	0.16	2.47	3.31	4.09*	0.78	0.63	5.40*	0.02	0.13	10.43**	21.66***

* $p < .05$; ** $p < .01$; *** $p < .001$

Critical $F_{1,130}$ = 3.92 at the 0.05 level of confidence

The number of degrees of freedom differs for each criterion variable: df = 130 is the df associated with the full regression model, and is the lower bound for the interim models against which the main effects are tested.

TABLE 9.1.2

PERCENTAGE OF VARIANCE IN EACH CRITERION VARIABLE EXPLAINED BY VARIOUS REGRESSION MODELS AND INDIVIDUAL PREDICTORS.

Criterion	R^2[a] Full Model	R^2 change							R^2[b] Revised Model	R^2 change							R^2change[c] I	R^2[d] Final Model	$\hat{R}^2$[e]
		T x PF1	T x PF2	T x PF3	T x PF4	T x CP1	T x CP2	T x I		PF1	PF2	PF3	PF4	CP1	CP2	T			
VABL	19.9	1.3	0.0	0.4	1.1	0.9	1.0	0.8	15.4	0.0	0.8	3.8	5.1	0.0	0.4	0.1	4.9	14.3	12.0
VABP	23.8	0.3	0.0	0.3	2.8	3.0	0.1	0.1	23.1	0.0	0.0	12.6	0.0	0.9	0.0	3.1	4.9	22.1	19.3
VAS	12.3	0.5	2.3	0.0	0.0	0.2	1.0	0.2	8.2	0.2	0.0	2.1	0.6	0.2	1.6	3.0	0.5	3.3	2.3
EOB	18.9	0.0	0.1	0.1	0.8	2.4	0.0	0.0	15.5	3.4	1.5	7.7	0.4	0.9	0.3	0.3	0.5	11.8	10.5
AD1	37.2	0.0	0.1	0.1	2.6	1.1	4.0	0.0	35.9	0.7	0.1	2.1	4.4	0.0	0.1	2.7	12.0	34.9	32.1
AI1	32.8	0.0	0.9	2.1	0.4	0.1	1.3	1.7	28.4	2.2	0.4	0.3	2.8	0.0	0.1	5.5	11.5	27.6	25.0

[a] Full model contains 15 terms: treatment, 4 personality factor scores, 2 cognitive preference scores, general ability, and 7 attribute-treatment interaction terms.

[b] Revised model for each criterion variable contains all main effects terms and all significant interaction terms.

[c] Compared against a model containing significant interactions, treatment and I only.

[d] Final model for each criterion variable contains only terms which make a significant contribution to explanation of variance: main effects and interactions.

[e] Estimated squared multiple correlation in the population corrected for shrinkage.

criteria where the prediction equations contain main effects only. This observation tends to support the contention advanced in Chapter 2 that studies of main effects alone are too naively conceived to shed much light on classroom performance. It is also important to note that the three outcome variables on which interactions were found are the three most highly rated outcomes of the teachers participating in this study. As a post-experimental exercise, the eight teachers were asked to rank the four attitude scales and the two achievement tests in terms of importance in their own teaching priorities. It is, of course, quite possible that these teachers may have tried to create a favourable impression by conforming to their interpretations of priorities in the Web of Life course materials. Be that as it may, the ranks given are seen in Table 9.1.3.

TABLE 9.1.3

RANKING OF CRITERION VARIABLES BY PARTICIPATING TEACHERS

Variable	Teacher								Av.
	1	2	3	4	5	6	7	8	
VABL	4	4	5	4	5	4	3	4	4.1
VABP	1	1	2	2	2	1	2	1	1.5
VAS	6	6	6	6	6	6	6	6	6.0
EOB	5	5	3	5	4	5	5	3	4.4
AD1	3	3	4	3	3	3	4	5	3.5
AI1	2	2	1	1	1	2	1	2	1.5

If VABP, AD1, and AI1 are seen by the teachers to be the most important of these outcomes, it is reasonable to assume that they devoted most of their teaching effort in these directions, thus giving greater possibility to the interaction of teaching style and student attributes. There is a main effect for treatment only with VAS which, however, only accounts for 3.3% of the observed variance of VAS scores, and no treatment effect at all with VABL and EOB.

The two cognitive preference variables, for which high hopes were held, contributed two significant interaction effects, and one which just failed to reach statistical significance ($F = 3.85$, critical $F(1,130) = 3.92$ at $\alpha = 0.05$ level of confidence).

Of the seventeen significant main effects, four were due to treatment differences, four to intelligence, and four to intellectual disposition (PF3). Of the other five, three related to a liberal, non-dogmatic outlook (PF4) and two to the cautious, convergent outlook (PF1).

The investigation found no main effects, nor interaction effects with treatment, for personality factor 2 which reflects levels of psychological adjustment and anxiety. It may be that the inquiry-based strategy of Web of Life, and the test situations associated with it, are not as stressful to some students as may have been expected.

The practical significance of the relationship of the predictor variables to each criterion variable can best be gauged from an inspection of Table 9.1.2. Multiple correlation can be regarded as the correlation between scores predicted by the regression equation and the observed scores on the criterion variable. The R^2 values indicate the percentage variance in each criterion variable explained by the terms in the regression equation acting jointly, and the R^2 change values reflect the loss of explanatory power accompanying the removal of the unique effect of that particular variable from the regression model. Because of the manner in which the unique contributions of the variables were defined, the R^2 change associated with intelligence represents the total contribution of intelligence to explanation of variance. $\hat{R}^2$ is the estimated squared multiple correlation in the population corrected for chance over-fit, or what is referred to by Kerlinger and Pedhazur (1973) as 'shrinkage'. As Kerlinger and Pedhazur explain:

> In calculating the weights to obtain a maximum R, the zero-order correlations are treated as if they were error-free. This of course is never the case. Consequently there is a certain amount of capitalization on chance, and the resulting R is biased upwards. (p.282)

The amount of shrinkage can be estimated by

$$\hat{R}^2 = 1 - (1-R^2)\ (N-1/N-1-V)$$

where $\hat{R}^2$ is the estimated squared multiple correlation in the population; R^2 is the obtained squared multiple correlation; N is the size of the sample; and V is the number of independent variables. It can be seen that for a given number of independent variables the greater the size of the sample the smaller the shrinkage, or as in this study with a fixed sample size, the fewer variables in the final regression equation for each criterion the smaller the shrinkage.

Using these population estimates it may be claimed that the final regression equations for VABP, AD1, and AI1 (which might be regarded as the emphasized outcomes of the Web of Life biology course), as prediction equations, are of considerable practical significance: even substantial on the knowledge-oriented achievement test AD1. On criteria VABL and EOB they are of moderate importance, and for VAS the prediction equation is of little practical significance.

9.2 STATISTICAL INFERENCE

This investigation is predominantly concerned with describing the relationships between instructional treatments and student attributes as measured in affective and cognitive outcomes. However, it is reasonable to ask "with what confidence can these relationships be generalized to a target population having similar characteristics to the sample population?"

To test the null hypothesis that there is no linear relationship between each dependent variable and the set of independent variables contained in the final regression equation (1-6 above) for that variable, is to test that the multiple correlation is zero in the population from which the sample is drawn.

The statistic used for the overall test is the F ratio, which can be defined either in terms of sums of squares or R^2, the squared multiple correlation coefficient:

$$F = \frac{R^2/V}{(1-R^2)/(N-1-V)}$$

where N is the sample size, and V the number of independent variables in the equation. Table 9.2.1 gives the multiple R, R^2, and overall F ratio for the final regression equation for each of the dependent variables.

TABLE 9.2.1

MULTIPLE R, R^2, and F VALUE FOR THE FINAL REGRESSION EQUATION FOR SIX CRITERION VARIABLES.

Criterion	Multiple R	R^2	$\hat{R}^2$	df	F
VABL	0.378	0.143	0.120	3,142	7.89***
VABP	0.471	0.221	0.193	5,140	7.96***
VAS	0.182	0.033	0.023	1,144	4.93*
EOB	0.343	0.117	0.105	2,143	9.52***
AD1	0.591	0.349	0.321	6,139	12.43***
AI1	0.523	0.276	0.250	5,140	10.65***

* $p < .05$; ** $p < .01$; *** $p < .001$

For each of these criterion variables the null hypothesis is rejected. At this stage of the analysis this testing of the null hypothesis is confirmatory. For each criterion variable the full regression model was initially shown to have a significant non-zero multiple correlation, and throughout the analysis only those terms which did not significantly contribute to the explanatory power of the equation were dropped out.

The overall null hypothesis that the multiple correlation is zero is equivalent to the null hypothesis that all V regression coefficients are equal in the population (Kim & Kohout, 1975). To reject the null hypothesis is to accept that at least one of the regression coefficients has a non-zero value. In the multiple regression procedure described in Chapter 8, the decrement of R^2 (the explained sums of squares) due to the removal in turn of each independent variable from the full model was calculated, and only those variables whose contribution to the explained variance was significant at the $\alpha = 0.05$ level of confidence were retained. So that the null hypothesis, that each variable in the final equation has no linear effect on the dependent variable once the effects of the other variables is controlled for, can also be rejected.

9.3 THE FINAL TREND EQUATIONS

In the determination of the reduced models for each criterion variable the independent variables were redefined as described in Chapter 8. However, in determining the final trend equations, the factor scores for the personality factors and the raw scores for the cognitive preference scales and for general ability were used. For descriptive and predictive purposes these original variables have more meaning than the redefined variables and, in fact, differ little from them. It was believed that there may have been non-zero correlations between the attribute variables, and the purpose of redefinition was to achieve orthogonality. Table 9.3.1 shows the extent of the intercorrelation between the original attribute variables.

TABLE 9.3.1

INTERCORRELATIONS BETWEEN ATTRIBUTE VARIABLES

Variables	PF2	PF3	PF4	CP1	CP2	I
PF1	-0.006	-0.002	0.008	-0.090	0.009	0.134
PF2		-0.007	-0.012	0.088	-0.134	-0.003
PF3			-0.005	0.096	0.184*	-0.306**
PF4				0.087	-0.109	-0.013
CP1					0.013	-0.069
CP2						-0.143

* $p < .05$; ** $p < .01$

Critical r (145) = 0.167 at 0.05 level of confidence.

It can be seen that there are only two significant non-zero correlations: those of PF3 with CP2 and with I. These relationships are illustrated in Figure 9.3.1. All other intercorrelations are non-significant.

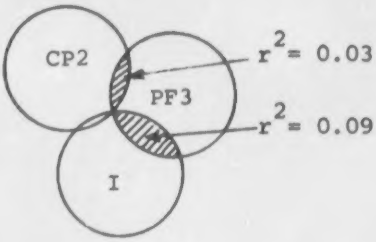


Figure 9.3.2 Intercorrelations between attribute variables

The redefinition of variables for the progressive reduction of the full regression model included the total contribution of general ability (which remains unaltered), and the unique contribution of the other variables. So that the only significant discrepancies between the original and redefined variables are that the unique contributions of CP2 and PF3 differ slightly from their total contributions.

9.4 VIEWS ABOUT BIOLOGY LEARNING (VABL)

For the criterion variable VABL, Table 9.4.1 gives the values of multiple R, R^2, and standard error of estimate, as well as the partial regression coefficients and β weights from multiple regression.

TABLE 9.4.1

SELECTED STATISTICS FROM MULTIPLE REGRESSION FOR CRITERION VARIABLE VABL

Multiple R	0.378
R^2	0.143
Standard error of estimate	5.183

Variable	b	β
PF3	-1.198	-0.217
PF4	-1.228	-0.223
I	0.107	0.157
Constant	31.816	

It can be seen from Table 9.4.1 that 14.3% of the variation in scores on VABL can be explained by PF3, PF4, and I acting jointly.

To predict a student's score on VABL the values of the constant, and the partial regression coefficients are used to obtain the prediction equation.

$$Y'_{VABLj} = 31.82 - 1.20\ PF3_j - 1.23\ PF4_j + 0.11I_j$$

The standard error of estimate is 5.18 so it is expected that on the average the predicted VABL score will deviate from the actual score by 5.18 units on the VABL scale (Kim & Kohout, 1973, p.331). However, the personality factor scores and the general ability score are in different units and it is difficult to judge the respective influence of each on VABL controlling for the others. The β weights or standardized regression coefficients can be used to make this comparison.

$$Z'_{VABLj} = -0.22Z_{PF3j} - 0.22Z_{PF4j} + 0.16Z_{Ij}$$

Other things being equal, one standard deviation unit increase in either PF3 or PF4 would result in a 0.22 z-score decrease in VABL, and a z-score unit increase in general ability would result in a 0.16 z-score increase in VABL.

An individual who scores highly on VABL tends to view biology learning as a non-authoritarian situation in which students are stimulated to think about biological phenomena, encouraged to discover biological relationships for themselves and to participate in the development of methods for solving these problems. If PF3 and PF4 are described in terms of the characteristics of low scorers on these factors, the relationship is simply expressed as follows: intelligence, a disposition to reflective and scholarly pursuits, and an independence of traditional beliefs all contribute to high scores on the VABL scale. There is no evidence that cognitive preferences or teaching style have had any significant influence; nor have any attribute-treatment interactions been demonstrated.

Little of the literature reviewed in earlier chapters refers to attitudinal outcomes. Gardner (1972), however, used essentially the same four scales as used in this investigation as dependent variables in a study of the attitudes of 11th grade physics students. For physics students he found that "the student who enjoys active participation in learning physics is the one who is serious, strives for success and enjoys intellectual activity, and also warmer in personal relationships" (p.347). There is little correspondence between the attribute measures used in Gardner's study and this one, but both find an intellectual disposition an important determinant of scores on VABL. Nor should it come as any surprise to find that intelligence and an independent, open belief system contribute to a liking for learning in an inquiry-oriented school subject.

9.5 VIEWS ABOUT BIOLOGY AS A PROCESS (VABP)

For the criterion variable VABP Table 9.5.1 gives the values of multiple R, R^2, and the standard error of estimate, as well as the partial regression coefficients and β weights from multiple regression.

TABLE 9.5.1

SELECTED STATISTICS FROM MULTIPLE REGRESSION
FOR CRITERION VARIABLE VABP

Multiple R	0.471
R^2	0.221
Standard error of estimate	3.505

Variable	b	β
Treatment	3.977	1.019
PF3	-1.301	-0.335
I	0.063	0.132
T x PF4	0.651	0.168
T x CP1	-0.191	-0.986
Constant	37.362	

It can be seen that 22.1% of the variation of scores on the VABP scale can be explained by T, PF3, I, T x PF4, and T x CP1 acting jointly.

Where instructional treatment is shown to have a main effect and/or interactive effects the contrast coding of 1, -1, provides two regression equations; one for each treatment. The prediction equations become, for treatment 1 (the individual-centred teaching style),

$$Y'_{VABPj} = 41.34 - 1.30\ PF3_j + 0.06\ I_j + 0.65\ PF4_j - 0.19\ CP1_j$$

and for treatment 2 (the class-centred teaching style),

$$Y'_{VABPj} = 33.39 - 1.30\ PF3_j + 0.06 I_j - 0.65\ PF4_j + 0.19\ CP1_j$$

both with a standard error of estimate of 3.51.

To gauge the respective influence of each variable on the VABP scores, controlling for all others, the appropriate z-score formulae are:

for T1, $Z'_{VABPj} = 1.02 - 0.33Z_{PF3j} + 0.13Z_{Ij} + 0.17Z_{PF4j} - 0.99Z_{CP1j}$

for T2, $Z'_{VABPj} = -1.02 - 0.33Z_{PF3j} + 0.13Z_{Ij} - 0.17Z_{PF4j} + 0.99Z_{CP1j}$

An individual who scores highly on the VABP scale tends to view biology as an open rather than a closed process which by it's very nature is dynamic, creative, tentative and unfinished. Intelligence and an intellectual disposition contribute to VABP scores, but in addition there are differential treatment and attribute-treatment interactions. In the prediction equations the treatment main effect becomes incorporated in the constant and the interaction effects are shown by the differences in sign between the treatments. The nature of the interactions can be better appreciated in the graphical representations in Figures 9.5.2, 9.5.3, and 9.5.4.

In plotting the values of attribute variables for the contrasting treatments against criterion scores, scores at, or approximately at, one standard deviation above the mean and one standard deviation below the mean have been used. This will give a consistent meaning to the terms 'high' and 'low' in the ensuing discussion.

In Figure 9.5.2, PF4 scores are plotted against VABP scores for treatments 1 and 2, at the mean values of PF3, I, and CP1. Similarly in Figure 9.5.3, CP1 scores are plotted against VABP scores for treatments 1 and 2, at the mean values of PF3, PF4, and I.

In Figure 9.5.2 a disordinal interaction between treatment and PF4 is seen: at the mean values on the other attribute variables the individual-centred teaching style is better than the class-centred style for the student who is dogmatic, authority-centred in outlook, and who tends to be gregarious, whereas the class-centred style is better for the student with an open outlook, sceptical of authority, independent of tradition, and with a preference for working alone.

Figure 9.5.3 illustrates a disordinal interaction between CP1 and treatment. At the mean values on the other attribute variables, the individual-centred teaching style is better than the class-centred style for students with a cognitive preference for practical applications, whereas the class-centred style is better for students whose cognitive preference is for a theoretical orientation.

In Figure 9.5.4 both PF4 scores and CP1 scores are plotted against VABP scores for treatments 1 and 2, at the mean values for I and PF3, to provide a three-dimensional representation.

It can be appreciated that the main effects of I and CP3 would merely move the intersecting regression lines in Figures 9.5.2 and 9.5.3 parallel to the vertical axis without altering the abcissa of the intersection. However it can be seen from Figure 9.5.4 that an alteration in either PF4 or CP1, holding the other constant, will result in moving the intersection of the regression lines along the intersection of the planes representing treatment 1 and 2. It is important to recognize the interrelatedness of these two interactions when attempting to predict the comparative superiority of treatment 1 or treatment 2

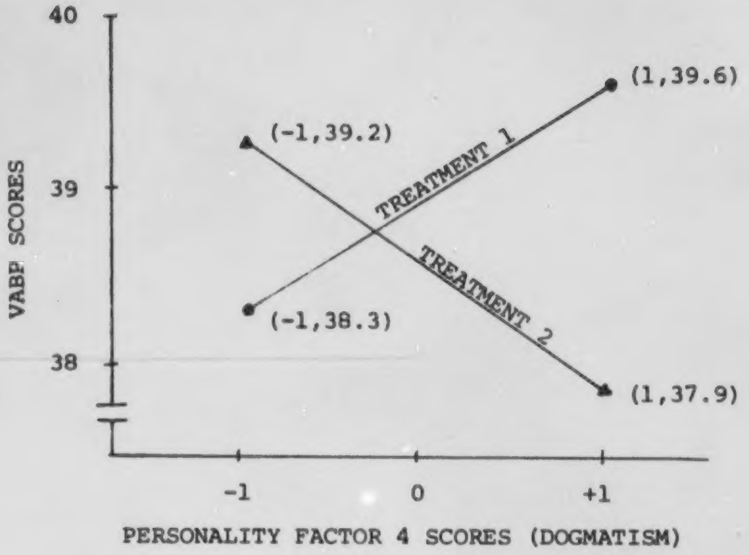


Figure 9.5.2 Plot of personality factor 4 (Dogmatism) scores against VABP scores showing the PF4 x treatment interaction

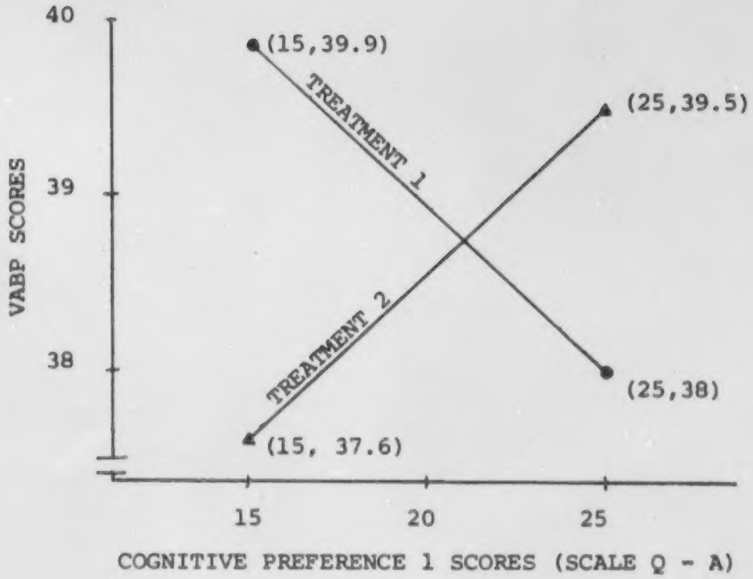


Figure 9.5.3 Plot of cognitive preference scale Q - A scores against VABP scores showing the CP1 x treatment interaction

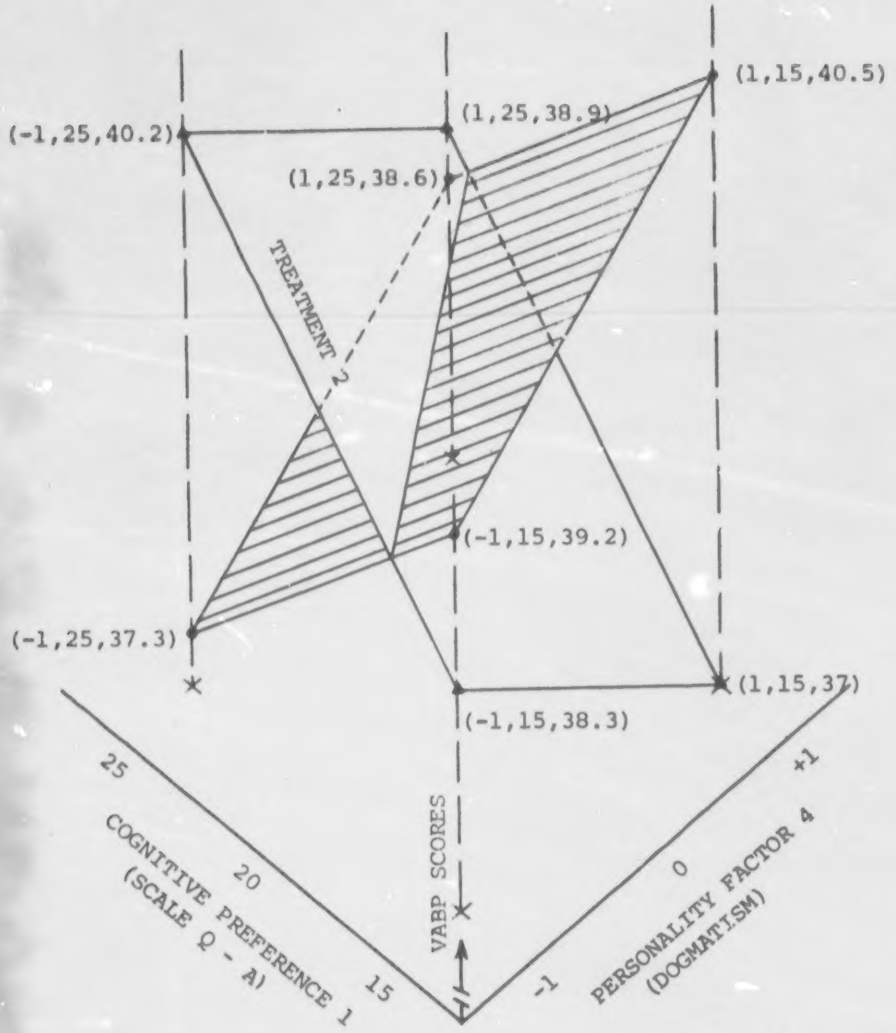


Figure 9.5.4. Combined plot of personality factor 4 (Dogmatism) scores and cognitive preference 1 (Scale Q - A) scores against VABP scores.

for an individual student. This can be illustrated by examples. Figure 9.5.2 indicates that, at the mean values of the other variables in the regression equation, treatment 1 is better than treatment 2 for students with a PF4 score of +1. Treatment 1 remains better for these students for all values of I and of PF3, but not for all values of CP1. As can be seen in Figure 9.5.4, treatment 2 is better than treatment 1 for these high dogmatic students whose cognitive preference score on scale Q-A is 24.2 or higher. Similarly, for students with a PF4 score of -1, treatment 2 is better than treatment 1 for all students whose CP1 score is 17.4 or better; for those whose score is below 17.4, treatment 1 is better.

There is no simple interpretation of these data because of the interrelatedness of the interactions. For students who are high on the dogmatism factor and have a strong preference for applied rather than theoretical presentations of biological material, the individual-centred, questioning teaching style appears to stimulate them to view biology as an open and tentative process. For those who have a strongly open-minded outlook and a cognitive preference for biological theory and abstractions, the class-centred teaching style appears more appropriate. These results may be related to the suggestion by Cronbach and Snow (1969) that treatments that benefit students who are deficient in some particular aptitude do so by providing information-processing functions they do not possess themselves. The dogmatic personality and the applied cognitive preference may need not only the additional support given by individual-centred teaching, but also the increased questioning provided by this style. The open-minded student with a theoretical cognitive preference may then perform better where he has to do more for himself. Decreased performance on the part of an able group under conditions of high instructional support has led Tobias (1976) and Cronbach and Snow (1969) to suggest that boredom and lack of attentiveness may be responsible.

9.6 VIEWS ABOUT SCIENTISTS (VAS)

For the criterion variable VAS, Table 9.6.1 gives the values for multiple R, R^2, and the standard error of estimate, as well as the partial regression coefficients and β weights from multiple regression.

TABLE 9.6.1

SELECTED STATISTICS FROM MULTIPLE REGRESSION
FOR CRITERION VARIABLE VAS

Multiple R	0.182	
R^2	0.033	
Standard error of estimate	4.401	
Variable	b	β
Treatment	0.18	0.18
Constant	36.46	

It can be seen that 3.3% of the variation of scores on VAS can be explained by the effect of treatment, whereas no attribute variable or interaction term contributed significantly to the explained variance.

The best prediction of an individual student's score on VAS is 37.27 for those with teachers with an individual-centred teaching style and 35.65 for those with teachers with a class-centred style, with a standard error of estimate of 4.40.

A student who scores highly on VAS tends to see biologists as normal, active, occasionally fallible human beings who are different from other people only in the area of their special training.

The full regression model explains 12.3% of the variance of scores on VAS (Table 9.1.2), and the only independent variable whose contribution was statistically significant was treatment, which explains only 3.3%. This suggests that a number of other variables made minor contributions which, nevertheless, did not reach the critical F ratio at the $\alpha = 0.05$ level of significance: variables such as T x PF2 (R^2 change = .023), CP2 (R^2 change = .016), and PF3 (R^2 change = .023).

A prediction equation which accounts for only 3.3% of the variance of scores on a particular criterion is of little practical significance. It is clear that the principal contribution is made by factors not considered in this research; maybe by factors such as parental and social influences. Within the boundaries of the present research, however, it is not surprising that students in classes where teachers have more personal contact with individuals tend to view biologists more as normal human beings than those in classes where the principal interaction between teacher and students is at the level of the class-as-a-whole.

9.7 ENJOYMENT OF BIOLOGY (EOB)

For the criterion variable EOB Table 9.7.1 gives the values for multiple R, R^2, and the standard error of estimate, as well as the partial regression coefficients and β weights from multiple regression.

It can be seen that 11.7% of the variation of scores on EOB can be explained by PF1 and PF3 acting jointly.

The prediction equation becomes

$$Y'_{EOBj} = 31.16 - 1.20\ PF1_j - 1.84\ PF3_j$$

To gauge the respective influence of each variable on the EOB scores, controlling for the other variable, the standardized regression coefficients, of β weights may be used:

$$Z'_{EOBj} = -0.19Z_{PF1j} - 0.29Z_{PF3_j}$$

TABLE 9.7.1

SELECTED STATISTICS FROM MULTIPLE REGRESSION FOR CRITERION VARIABLE EOB

Multiple R	0.343
R^2	0.117
Standard error of estimate	6.085

Variable	b	β
PF1	-1.201	-0.186
PF3	-1.844	-0.288
Constant	36.164	

Other things being equal, one z-score unit increase in PF1 would result in a 0.19 unit decrease in EOB, and a unit increase in PF3 would result in a 0.29 unit decrease in EOB.

An individual who scores highly on EOB tends to view biology as an important activity for himself, and to enjoy actively searching for an understanding of biological phenomena. If PF1 and PF3 are described in terms of the characteristics of low scorers on these factors, the relationship is simply expressed as follows: a cautious, convergent personality and an intellectual disposition both contribute to high scores on the EOB scale, particularly the latter. There is no evidence that general ability, cognitive preference, or teaching style have had any significant influence, nor have any significant attribute-treatment interactions been demonstrated on this criterion.

9.8 ACHIEVEMENT TEST AD1 (DIVERSITY OF LIVING ORGANISMS)

For the criterion variable AD1 Table 9.8.1 gives the values of multiple R, R^2, and the standard error of estimate, as well as the partial regression coefficients and β weights from multiple regression.

TABLE 9.8.1

SELECTED STATISTICS FROM MULTIPLE REGRESSION FOR CRITERION VARIABLE AD1

Multiple R	0.591
R^2	0.349
Standard error of estimate	3.247

Variable	b	β
Treatment	-4.213	-1.069
PF3	-0.596	-0.152
PF4	-0.871	-0.222
I	0.150	0.309
T x PF4	0.628	0.160
T x CP2	0.272	1.230
Constant	7.701	

It can be seen that 34.9% of the variation in scores on AD1 can be explained by the joint linear effect of T, PF3, PF4, I, T x PF4, and T x CP2.

When instructional treatment is shown to have a main effect and/or interaction effects, the contrast coding of 1, -1, produces two prediction equations: one for each treatment. The main effect of treatment becomes incorporated into the constant and the interaction effects are shown by the differences in sign and the magnitudes of the coefficients between the treatments. In the reduced regression model for AD1, PF4 appears both as a main effect and in an interaction term. It is appropriate to consider both effects together, and the substitution of the partial regression coefficients in the reduced model formula accomplishes this. The prediction equations become, for treatment 1 (individual-centred teaching style),

$$Y'_{AD1j} = 3.49 - 0.60\ PF3_j - 0.24\ PF4_j + 0.15\ I_j + 0.27\ CP2_j$$

and for treatment 2 (class-centred teaching style),

$$Y'_{AD1j} = 11.91 - 0.60\ PF3_j - 1.50\ PF4_j + 0.15\ I_j - 0.27\ CP2_j$$

both with a standard error of estimate of 3.25 units on the AD1 test.

To gauge the respective influences of each variable on the AD1 scores, controlling for all other variables, the appropriate standard-score formulae are, for treatment 1,

$$Z'_{AD1j} = -1.07 - 0.15Z_{PF3j} - 0.06Z_{PF4j} + 0.31Z_{Ij} + 1.23Z_{CP2j}$$

and for treatment 2,

$$Z'_{AD1j} = 1.07 - 0.15Z_{PF3j} - 0.38Z_{PF4j} + 0.31Z_{Ij} - 1.23Z_{CP2j}$$

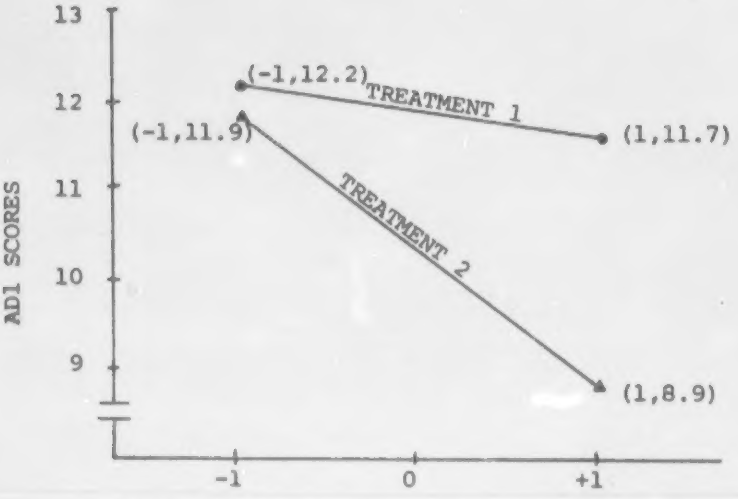
Other things being equal, one standard deviation unit increase in general ability would result in a 0.31 z-score increase in AD1, and a standard deviation unit increase in PF3 would result in a 0.15 unit decrease in AD1. There are, however, two significant attribute-treatment interaction effects. The effect of the T x PF4 interaction modifies the main effect of PF4, so that for treatment 1 one standard deviation unit increase in PF4 results in a decrease of 0.06 z-score units in AD1, whereas for treatment 2 the same increase results in a decrease of 0.38 z-score units in AD1. A standard deviation unit increase in CP2 would lead to a 1.23 unit increase in AD1 for treatment 1, and a 1.23 decrease in AD1 for treatment 2.

A student who scores well on test AD1 has a good knowledge of the classification of plants, animals, and microorganisms. The test, almost exclusively, calls for memory alone: only two items out of twenty-five call for any re-ordering or processing of remembered information. Both intelligence and an intellectual disposition contribute to the AD1 scores but in addition there are differential treatment and attribute-treatment interaction effects.

The nature of the interactions can be illustrated graphically. In Figure 9.8.2 PF4 scores are plotted against AD1 scores for treatments 1 and 2 at the mean values for the other variables in the equation. In Figure 9.8.3 CP2 scores are plotted in a similar fashion.

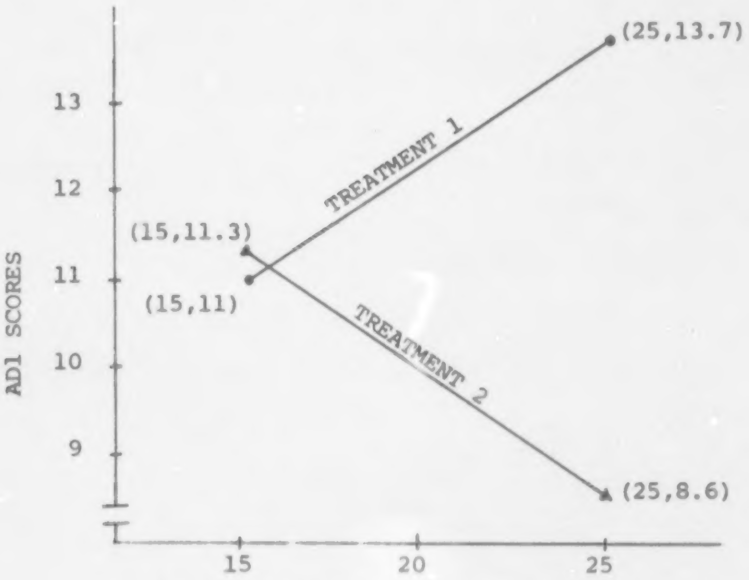
It can be seen from Figure 9.8.2 that the multiple regression solution giving a main effect for PF4 as well as a T x PF4 interaction has the net effect of producing regression lines for the alternative treatments which have gradients of the same sign but different magnitude; an effect which Bracht and Glass (1970) would regard as an ordinal interaction. At the mean values for the other variables, the individual-centred teaching style is better than the class-centred style for students with PF4 scores between -1 and +1. Under individual centred teaching open-minded students perform slightly better than dogmatic students, but under class-centred teaching open-minded students perform markedly better than dogmatic students. This finding is consistent with the research studies on dogmatism discussed in Chapter 5.

Figure 9.8.3 illustrates the interaction between CP2 and treatment. At the mean values of the other variables the individual-centred teaching style is markedly superior to the class-centred style for students with a cognitive preference for understanding principles rather than rote learning, whereas the class-centred style becomes the more effective treatment for students who have a strong preference for rote learning. The marked feature about this interaction is the effect the two teaching styles have on students with a preference for principles and generalizations. The individual, questioning style appears to stimulate them whereas the class-centred and more tightly



PERSONALITY FACTOR 4 SCORES (DOGMATISM)

Figure 9.8.2 Plot of personality factor 4 (Dogmatism) scores against AD1 scores showing the PF4 x treatment interaction



COGNITIVE PREFERENCE 2 SCORES (SCALE P - M)

Figure 9.8.3 Plot of cognitive preference scale P - M scores against AD1 scores showing the CP2 x treatment interaction

supervised teaching style appears to inhibit their performance.

The main effects of I and PF3 will merely move the regression lines in Figures 9.8.2 and 9.8.3 parallel to the vertical axis without altering their relationship to one another. However, when either PF4 or CP2 are altered, holding the other constant, the intersection of the regression lines moves along the intersection of the planes representing treatments 1 and 2. This relationship can be seen in Figure 9.8.4, where both PF4 scores and CP2 scores are plotted against AD1 scores for treatments 1 and 2 at the mean values of I and PF3.

The interrelatedness of the two interactions can be more readily appreciated by reference to Figure 9.8.4. Figure 9.8.2 indicates that at the mean values of the other variables in the regression equation treatment 1 is better than treatment 2 for students with a PF4 score of -1. Treatment 1 remains better for these students for all values of I and PF3, but not for all values of CP2. As can be seen from Figure 9.8.4, treatment 2 is better for students with a PF4 score of -1 whose score on the P-M scale is 17.8 or lower. And although, at the mean value for PF4, treatment 2 is better than treatment 1 for students with a CP2 score of 15, treatment 1 is better for those with a score on the P-M scale of 0.2 or higher. That is, for students who are both open-minded and have a cognitive preference for rote-memory, class-centred teaching leads to greater achievement on what is basically a memory task: for dogmatic students and those with a cognitive preference for principles and generalizations in biology, the individual-centred teaching style is better.

Perhaps closed-mindedness is a debilitating characteristic which needs the stimulus and the support given by an individual-centred teaching style, whereas the open-minded student with a preference for rote learning can do better on a memory task when allowed to exercise his cognitive preference without the 'interference' of additional

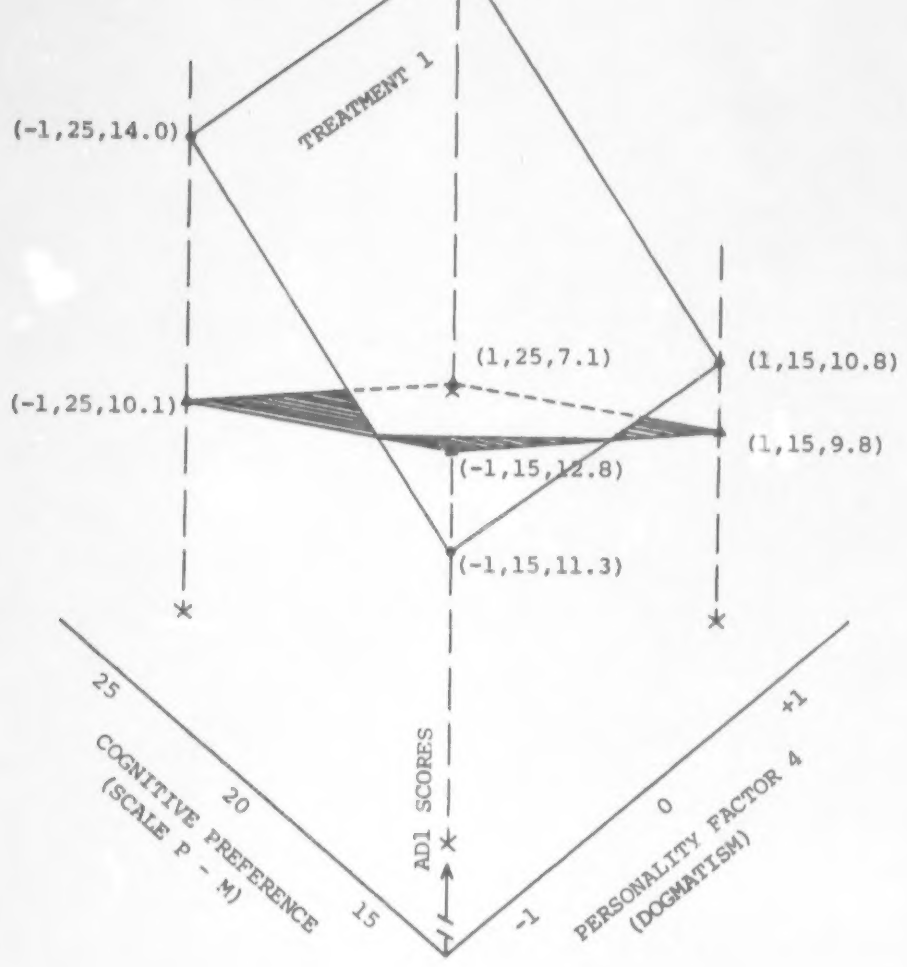


Figure 9.8.4. Combined plot of personality factor 4 (Dogmatism) scores and cognitive preference 2 (Scale P - M) scores against AD1 scores.

attention and stimulation. That is, the individual-centred teaching style, it is suggested, gives assistance to those who suffer some inability to accept the new information readily, and to those whose cognitive preference is not congruent with the learning task, but has an 'over-kill' effect on those who are able to accept the information readily and for whom the task is congruent with their cognitive preference.

9.9 ACHIEVEMENT TEST AI1 (INTER-RELATIONSHIPS)

For criterion variable AI1 Table 9.9.1 gives the values of multiple R, R^2, and the standard error of estimate, as well as the partial regression coefficients and β weights from multiple regression.

TABLE 9.9.1

SELECTED STATISTICS FROM MULTIPLE REGRESSION
FOR CRITERION TEST AI1

Multiple R	0.525
R^2	0.276
Standard error of estimate	2.610

Variable	b	β
Treatment	0.704	0.234
PF1	-0.453	-0.150
PF4	-0.502	-0.167
I	0.127	0.342
T x PF3	0.504	0.168
Constant	10.707	

It can be seen that 27.6% of the variation in the scores on AI1 can be explained by T, PF1, PF4, and T x PF3 acting jointly.

The prediction equations become, for treatment 1 (individual-centred teaching style),

$$Y'_{AI1j} = 11.41 - 0.45\ PF1_j + 0.50\ PF3_j - 0.50\ PF4_j + 0.13\ I_j$$

and for treatment 2 (class-centred teaching style),

$$Y'_{AI1j} = 10.01 - 0.45\ PF1_j - 0.50\ PF3_j - 0.50\ PF4_j + 0.13\ I_j$$

with a standard error of estimate of 2.61 units on the AI1 test.

To gauge the respective influence of each variable on the AI1 scores, controlling for all others, the appropriate standard-score formulae are, for treatment 1,

$$Z'_{AI1j} = 0.23 - 0.15Z_{PF1j} + 0.17Z_{PF3j} - 0.17Z_{PF4j} + 0.34Z_{Ij}$$

and for treatment 2,

$$Z'_{AI1j} = -0.23 - 0.15Z_{PF1j} - 0.17Z_{PF3j} - 0.17Z_{PF4j} + 0.34Z_{Ij}$$

Other things being equal, one standard deviation unit increase in I would result in a 0.34 unit increase in AI1, a unit increase in PF1 would result in a 0.15 decrease in AI1, a unit increase in PF4 would result in a 0.17 unit decrease in AI1. There is one significant interaction: a unit increase in PF3 would result in a 0.17 unit increase in AI1 under treatment 1, and a 0.17 unit decrease under treatment 2.

A student who scores well on test AI1 has a good knowledge of the inter-relationships between living organisms and can process this knowledge in a number of ways, either by reordering the information, by applying it to new or novel situations, or by using it to analyze the relationships between biological phenomena. Intelligence, open-mindedness, and a cautious, convergent personality all contribute to high scores on AI1, and there are differential treatment and attribute-treatment interaction effects which will be considered together.

The nature of the interaction is represented graphically in Figure

9.9.2, where the PF3 scores are plotted against AI1 scores for treatments 1 and 2, at the mean values for the other variables in the equation.

Within the range of PF3 scores -1 to +1, the individual-centred teaching style leads to higher achievement on AI1 than does the class-centred style, and is markedly superior for the student with the non-intellectual disposition. The individual assistance and questioning may provide the stimulus for intellectual activity that these students normally do not have. On the other hand, the class-centred teaching style appears to be counter-productive with these students.

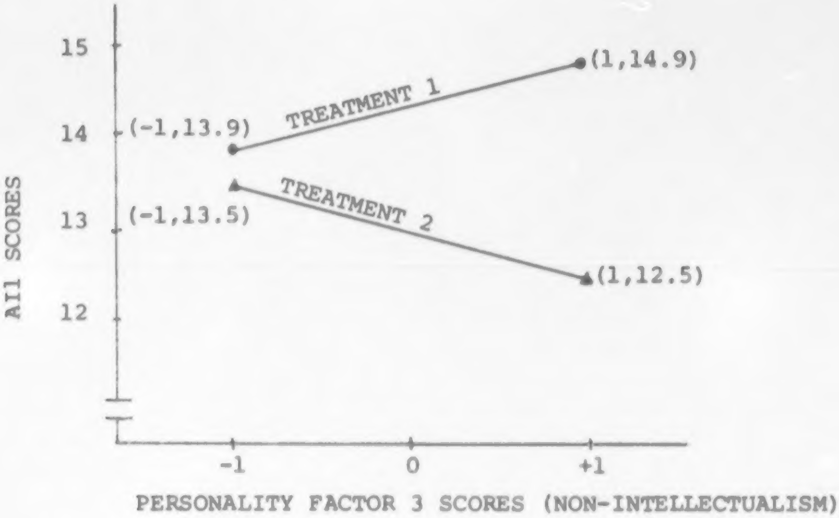


Figure 9.9.2 Plot of personality factor 3 (Non-intellectualism) scores against AI1 scores showing the PF3 x treatment interaction

If extrapolated, the regression lines in Figure 9.9.2 would intersect at (-1.4, 13.7), so that only at high levels of intellectual disposition (below -1.4) would the class-centred teaching style be more effective than the individual-centred style. Tobias (1976) and Cronbach and Snow (1969) have both noted indications in the research literature that high instructional support appears to be counter-productive with able students. They suggest that boredom and inattentiveness may be the cause.

The effects of intelligence, impulsivity, and dogmatism are additive and shift the regression lines up or down, parallel with the vertical axis, without altering their relationship to one another.

CHAPTER 10

IMPLICATIONS OF THE RESEARCH

10.1 SUMMARY OF RESULTS

10.2 IMPLICATIONS

10.3 SUGGESTIONS FOR FURTHER RESEARCH

10.4 CONCLUDING OBSERVATIONS

10.1 SUMMARY OF RESULTS

The results are described in detail in Chapter 9, but for the discussion which follows it will be useful to present a summary of the significant findings. Attribute-treatment interactions are not common, but they occur for the three criterion variables regarded by the participating teachers as the most important outcomes (Views about biology as a process, Achievement test AD1, and Achievement test AI1). It is for these dependent variables, also, that the final regression models explain the greatest amount of total variance.

The final trend equations give the most parsimonious description of the independent contribution of each of the fifteen predictor variables to explained variance of each criterion variable. These relationships are summarized:

1. Increased levels of general ability, a disposition to intellectual pursuits, and an independence of traditional beliefs, all contribute to a student viewing biology learning as a non-authoritarian and participatory activity.
2. Increased levels of general ability, and an intellectual disposition, contribute to a student viewing biology as an open, dynamic, and creative process. In addition, there are complex attribute-treatment interactions: in brief, for dogmatic students with a cognitive preference for application, the individual-centred teaching style appears to stimulate a view of biology as an open process, but it appears to be counter-productive for open-minded students with a cognitive preference for theory. For these latter students the class-centred teaching style appears to be better, although it is relatively ineffective for the dogmatic students with an 'applied' cognitive preference.
3. When corrected for chance over-fit, only 2.3% of the total variance in scores on 'Views about scientists' is explained by the final

regression equation: a relationship of little practical significance. The only variable considered in this investigation which significantly contributes to a student viewing scientists as normal human beings, different from others only in the area of their special training, is their experience of individual-centred rather than class-centred teaching.

4. Enjoyment of biology appears to be basically a function of two personality characteristics: a cautious convergent outlook, and an intellectual disposition.
5. Increased levels of general ability, and an intellectual disposition contribute to higher scores on an achievement test which is basically a memory task. There are also two complex attribute-treatment interactions which can be briefly described as follows: for students who have both a cognitive preference for rote learning and who are open-minded, class-centred teaching appears to lead to higher scores, but for dogmatic students and those with a cognitive preference for principles and generalizations, individual-centred teaching appears to be better.
6. Increased levels of general ability, open-mindedness, and a cautious convergent personality, all contribute to increased scores on an achievement test which requires a student to process his knowledge by reordering it, applying it to new situations, or analyzing its elements. Over the range of scores on non-intellectualism -1σ to $+1\sigma$, the individual-centred teaching style leads to higher achievement, and it is distinctly superior to class-centred teaching for the extreme non-intellectual student. For students with a highly intellectual outlook class-centred teaching becomes more effective, and individual-centred teaching becomes counter-productive.

10.2 IMPLICATIONS

Some disclaimers

With regard to the internal validity of the investigation, an attempt was made to secure co-educational schools set in a similar socio-economic environment and with similar physical provisions. Although intact class groups were used, and not randomly assigned to the alternative treatments, it has been argued that the allocation of individuals to treatments was random in effect. The preservation of the natural school setting and the natural assemblages of teachers and students, the use of the teacher's natural teaching styles over a full year's work, and the care taken to minimize experimental intrusion, were all attempts to preserve the ecological integrity of the setting. Ecological validity was valued ahead of control over extraneous independent variables. However, in generalizing to a wider population, the possibility of confounding between such variables and the experimental variables must be borne in mind. As to treatment differences, all teachers were teaching the same course content, and accepting the same rationale, objectives, and general teaching strategies described in the Web of Life curriculum materials: they were all recognized as good teachers who differed in their personal style of teaching.

It is argued that, within the adequacy of the instruments employed, differences in outcome for individual students can be ascribed to student attributes, treatment differences, and/or attribute-treatment interaction effects. The proportion of the total variance explained by the model reflects on its adequacy as a predictor.

Generalization, on the other hand, is a more hazardous undertaking, and the following cautions are in order. For purposes of generalization, the care taken to maximize the ecological validity of the investigation is offset to some extent by the use of an extreme-groups design. The participating teachers were selected on certain criteria and all who were approached agreed to take part in the investigation. They were volunteers, but only in the sense that every participant in a research project is a volunteer: non-volunteers do not participate. Nevertheless, the teachers selected were believed to possess certain teaching styles and can not be assumed to be representative of all biology teachers. They also adhered to the Web of Life prescriptions of the time, before the use of optional units and adaptations of the course were prevalent. The uncertainty, until the post-experimental analysis, as to which classes belonged to which treatment, provided protection against the Hawthorne effect, and other extraneous experimental effects, such as experimental intrusion and pretest effect, were not considered to affect adversely generalizability to non-experimental situations. So that, although generalizability to other natural school settings is enhanced by the design of the investigation, care should be exercised in generalizing to teachers or treatments unlike those used in the investigation.

The four assumptions on which parametric statistics are based are the independence of observations, interval scale data, normality of distribution, and homogeneity of variance (Siegel, 1956, p.19). In order to examine the effects of personological variables and of attribute-treatment interactions, the individual had to be used as the unit of analysis and so the first assumption is not strictly met. The robustness of tests of significance under violation of the remaining three assumptions is well attested (Cohen, 1968; Kerlinger & Pedhazur, 1973; Gardner, 1975d). Arguments of robustness, however, do not disguise the fact that the prior condition of random sampling, on which all statistics (parametric and non-parametric) are based, is not met in this, nor in most experimental research in education (see Chapter 3). For all that, it has been argued on Cornfield and Tukey's (1956) advice, that in order to interpret the results, statistical tests should be employed in generalizing to a hypothetical target population like that in the sample. However, some caution must be exercised in interpreting the levels of significance given. On the other hand, descriptive statistics such as the multiple correlation coefficient R, or the squared semi-partial correlations R^2, are valid for the sample, no matter how it was selected, and these statistics provide some indication of the practical significance of the relationships described.

Implications

If high scores on the criterion tests are measures of success, Chapter 9 provided the detailed answers to the question asked in Chapter 1: "For whom is the Web of Life course succeeding, and in the hands of which teachers?" In the most general terms, the course appears to be succeeding for steady (non-impulsive), intelligent, open-minded students with an intellectual disposition. Where treatment has an effect, an individual-centred, questioning teaching style appears to be more successful for closed-minded students (on criteria VABP and AD1); for students with a cognitive preference for application rather than theory (on criterion VABP); for students with a cognitive preference for principles rather than rote learning (on criterion AD1); and for non-intellectual students (on criterion AI1). On the other hand, a class-centred teaching style appears to be more successful for highly intellectual students (on criterion AI1), and for open-minded students (on criteria VABP and AD1), particularly those with a cognitive preference for theory rather than application (on criterion VABP) and those with a cognitive preference for rote learning rather than principles (on criterion AD1).

The attribute-treatment interactions found in this investigation provide still further evidence for the need for individualized instruction. The belief in individualized instruction is a belief in attribute-treatment interaction: that there is no one best way of learning which suits all students. As Salomon (1972) says:

> Those working in ATI research differ ... from those involved in traditional research in instruction in that they admit the existence of individual differences and take them into consideration. Their work is based on the assumption that different instructional procedures interacting with aptitudes will lead to the same learning outcomes. (p.327)

To maximize cognitive and affective outcomes, these data could be used as a basis for grouping students with teachers, where there are teachers available who exhibit these teaching styles (e.g., Cunningham, 1975; Bennett, 1976). The identification of teaching styles need not be so elaborate or time-consuming as in this investigation: now that there are indications that behaviours such as giving individual assistance and questioning are important, and that teacher behaviour is relatively stable over time, a relatively economical observation schedule could be used to differentiate between teachers on these characteristics.

Although in this investigation the characteristics of natural teaching styles were identified, that is not to say that these patterns of behaviour can not be adopted by others who do not possess them. By raising these behaviour patterns to the level of consciousness they may be redefined as deliberate elements of teaching strategy. Teachers, and particularly teachers-in-training, may be encouraged to practise the skills associated with giving individual assistance and stimulation within class groupings. Hunt (1974, Note 3) describes a number of skills required for meeting the needs of individual students: skills in distinguishing between students, learning environments and outcomes; skills in 'radiating' particular learning environments; and skills in 'flexible modulation' from one environment to another. In teacher training courses these skills could be practised in micro-teaching sessions with peer-group role players, or with school students selected because of their differences.

Stripped of the detail, the research findings could be interpreted as suggesting that teachers give the closed-minded student, the non-intellectual student, and the student with a cognitive preference for application, the individual assistance and stimulation they need, but to remember that this stimulation is counter-productive for open-minded, and intellectually disposed students. Salomon (1972) describes three heuristic models for ATI research: the remedial model, the compensatory model, and the preferential model. In the remedial model, learning is regarded as connective and cumulative, and when some essential task-specific element of knowledge is missing no progress can be made until the deficiency is made up. In the compensatory model deficiencies are viewed as relating to more general information-processing abilities, and their debilitating effects can be circumvented by providing the structure that the student can not provide for himself. A clear illustration of the preferential model can be seen in relation to cognitive preferences where a treatment may try to capitalize on a student's preferred information-processing strategy. Of the five attribute-treatment interactions found in this investigation, four can be described in terms of the compensatory model, and one in terms of the preferential model: the superiority of individual-centred teaching for students with a cognitive preference for principles, and of class-centred teaching for students with a cognitive preference for rote learning on a rote learning task.

When considering the implications of research, value judgements and positions play a large part. Perhaps teaching to promote maximum learning is not enough. It could be argued that to circumvent a student's deficiency in information-processing or to capitalize on his present capabilities is not in the student's best long-term interests. Perhaps teaching should be directed towards remedying deficiencies in information-processing and towards fostering alternative modes of thinking about things, otherwise the student may never learn to discover anything for himself (see Cronbach & Snow, 1969, p. 178; Messick, 1970). Mayer (1975) found that, where meaningful learning was the goal, instructional treatment interacted with students' prior experience, and says that his results

> need not be interpreted as suggesting that learners low in relevant experience should be taught by the less meaningful method; a more useful procedure ... is to provide such learners with the prerequisite knowledge they need and proceed with meaningful instruction. (p. 540)

Hunt's (1975) view that interaction should be seen in developmental perspective is of some relevance here: not only should a student's present needs for individual assistance be met, but his need should be viewed as a stage in his development towards a long-term objective of independence and less need for assistance.

The remedial model appears to be most clearly in the students' short and long-term interests, and there is obviously great promise in adapting instruction to individual differences with this principle of remediation in mind. Two important instructional strategies which meet this need are, firstly, the sequencing and ordering of content in various subject matter fields (see Posner & Strike, 1976), by developing and validating learning hierarchies (e.g., Gagné, 1968, 1970; Wang, Resnick, & Boozer, 1972; White, 1974), and secondly, the adoption of the mastery learning model (e.g., Carroll, 1963; Bloom, 1968; Block, 1971).

On the other hand, if maximizing achievement is the goal, perhaps the counselling model should be adopted, and the relationships between student attributes and achievement should be used to counsel students in their choice of courses. Should all non-intellectual and closed-minded students be counselled against taking biology? If so, for what courses, if any, would these characteristics be advantageous? Witkin et al. (1977) cautions against the use of relationships found to exist between certain measures of cognitive style and achievement "to perpetuate a self-fulfilling prophecy" (p. 53). This investigation holds out some promise to the non-intellectual and the closed-minded student that certain teaching styles are beneficial to them.

10.3 SUGGESTIONS FOR FURTHER RESEARCH

This investigation has been exploratory and descriptive, and has attempted to put into practice many of the suggestions made by earlier researchers and reviewers. Some of the suggestions required a 'trade-off' of one advantage for another. For example, increased ecological representativeness was traded off against reduced control over extraneous independent variables; an extreme-groups design to increase the prospects of discovering attribute-treatment interactions was traded off against random sampling. The generalizability of the study is limited largely by the extreme-groups design and by the limited sampling of teachers. With insights generated from this investigation, replication of the study could test its wider generalizability. By using more economical means of gathering data on classroom behaviour, it would be possible to use a larger, and random, sample of teachers. Such an extension would also necessitate consideration of other environmental variables such as socio-economic status, home-support for learning, community and peer-group pressures. The design of the study would also be improved by the inclusion of a measure of the students' perception of the teachers' styles (e.g., Tuckman, 1970) to check on phenomenological validity (Baird, 1973; Bronfenbrenner, 1976). If money and assistance were available it would be desirable to add to the criterion tests a good test of problem-solving skills in the laboratory (e.g., Tamir, 1974), thus covering the range of the stated objectives of the Web of Life course.

It would be useful to know, in the context of the Web of Life course, to what extent cognitive preferences are modifiable. Such data would add to the evidence on the stability of these variables, and would have implications for the anticipated outcomes of the Web of Life course.

The authors of the Web of Life course, in commenting on the affective objectives of the course, appreciate that students may come to the course with favourable attitudes (after all, it is not a compulsory subject, and is selected by students on the basis of interest or propaedeutic need) and hope that students "will have maintained or developed" appropriate interests and attitudes (Teacher's Guide, AAS, 1976, p.21). Consequently, in this investigation the possession of the appropriate attitudes was taken as the criterion, but it would be of interest to know how attitudes change throughout the course for different students with different teachers.

Naturalistic studies such as this, which make no special demands on, or modification to, normal school routines, can not hope to elucidate the fine detail of student response to teaching. That can only be done by an intensive case-study approach. On the other hand, there would appear to be a need to resort to more strictly controlled experimentation to validate the relationships suggested in this field study. In Chapter 4 a substantial argument was advanced against experimental studies of a brief and artificial nature because of their lack of ecological validity, and there are substantial practical difficulties in attempting to carry out controlled experimental studies in the classroom, quite apart from the ethical problem of holding some students for substantial periods of time in programs which are known not to be in their best interests. Nevertheless,there is a need for further investigation of information-processing variables and the ways in which they interact with content presentation and instructional strategy (see Mayer, 1975). Webster and Mendro (1974), impressed by Salomon's (1972) three models for ATI research, make a practitioner's plea for systematic task analysis of the skills associated with the mastering of basic educational objectives, as a precursor to the identification of treatments designed to interact with these skills in natural classroom settings. Hunt (1974, Note 3) and his co-workers are accumulating important data on the functional matching of characteristics of the instructional environment with the cognitive orientation of students.

There are other potentially fruitful analyses which could be performed on the data already collected in this investigation. The annotated transcripts of lessons and the accompanying audiotapes are a fertile field for analyses from different viewpoints, looking at, for example, sequences of behaviour such as the nature of student response to a teacher, or teacher response to students, or the logical nature of the discourse. Different characterisations of the teachers may well arise from such different analyses.

It has been shown that the relationships between student attributes and instructional treatment are not the same for different outcomes within the same subject area. Neither is it assumed that these relationships would be the same in other subjects, even in other science subjects which have a similar emphasis on inquiry. In fact, there are strong suggestions that attribute-treatment interactions may be specific to particular content (e.g., Nuthall, 1968; Tallmadge & Shearer, 1969). Replications in other subject fields would shed light on the worth of this line of investigation throughout the secondary school curriculum.

10.4 CONCLUDING OBSERVATIONS

The focus of this investigation has been on classrooms of one particular kind. Each classroom is a unique assemblage of teacher and students gathered together for one purpose: that the students may learn. The teacher brings to this enterprise his own peculiar blend of personal and professional competencies, as do the students. The interaction between them is infinitely rich in complexity.

In order to improve the conditions for learning, investigators of various persuasions have attempted to analyze this complex classroom interaction. Analysts search for order and for patterns, and can only do this by adopting a particular investigatory framework. The injunction to let the data speak for themselves can not, in fact, be taken literally: the data do not speak, their meaning and significance has to be discovered. However, the injunction urges the analyst to be aware that he is adopting a particular investigatory framework, and urges him to attempt to discover patterns in the data rather than to impose patterns on the data.

Non-quantitative methods such as case studies, participant observation, comparative and historical analyses all have their particular strengths and weaknesses in attempting to describe what has gone on in a classroom. The analytical framework adopted in this investigation has been a quantitative one and, within the limits of this approach, it has been possible to analyse large amounts of multivariate data. However, such a framework is not well suited to the analysis of the quality of classroom interaction as there are no readily accepted quantitative indices of quality, and so greater insight into classroom interactions will be gained by synthesizing the results from investigations with differing analytical frameworks. A renaissance man is required to perform the synthesis.

APPENDIX I

UNPUBLISHED INSTRUMENTS USED IN THE PRESENT RESEARCH

Modifications to administration of Omnibus Personality Inventory

Biology Preference Test

Biology Attitude Inventory

OMNIBUS PERSONALITY INVENTORY - FORM F

(modifications to administration)

This is a 385 true-false item multiphasic personality inventory developed by Heist, P., Yonge, G., Mc Connell, T.R., and Webster, H. and copyrighted and published by The Psychological Corporation, New York, N.Y. Answer sheets may be machine scored or hand scored to provide scores on 14 personality scales.

Because the inventory was developed in the USA, and standardized against a population of college freshmen, a trial administration of the OPI was carried out with one 11th grade biology class to check on any difficulties with language, and to note the time requirements. As a result of the trial administration the following additional instruction was given to students when administering the OPI.

> This test has been designed for use in American schools and some of the words and phrases may be strange to you. On the blackboard I have written down the common meanings of the more difficult ones. If you come to a word you don't understand you should check its meaning with the list on the board.

On the blackboard was written the following list.

No.	3.	cordial	friendly
	50.	esthetic	appreciation of beauty (in art, music, literature, drama, etc.)
	64.	profane	blasphemous swearing
	66.	novelty	something new or unusual
	84.	conventional	traditional, 'square'
	130.	radical	politically reformist, 'left-wing'
	161.	compromise	'giving ground' to
	178.	metaphors	e.g., using the phrase "to put under a microscope" to mean to examine in detail.
	178.	analogies	reasoning from similar cases

213.	tantalized	teased, kept curious
220.	'cutting up'	misbehaving
237.	primitive music	with simple instruments such as drums and pipes, emphasizing rhythm
238.	callous	hard, unfeeling
243.	agnostic	not knowing whether or not there is a God
245.	to be a booster for	to praise
293.	abstractionism	art which does not represent real objects
308.	worldliness	concerned with earthly pleasures
370.	fascism	politically extreme 'right-wing' and anti-communist

BIOLOGY PREFERENCE TEST

Time allowed: 15 minutes

DO NOT OPEN THIS BOOKLET UNTIL YOU ARE TOLD TO DO SO

This is not a factual test and there are no wrong answers. This test is designed to find out how students think and feel about different areas of biology.

Each item in this test begins with an introductory statement or diagram followed by four choices lettered A to D. Each of the choices is related to the introductory statement and each one is factually correct.

Read each item carefully. Select from the four choices the one which you find most appealing or satisfying, that is, the one which seems to you to be related to the introductory statement or diagram in the most appealing or most satisfying way. Write the letter of this choice in the appropriate space on the answer sheet.

Then select the choice which appeals to you least of all and write the letter of this choice in the appropriate space on the answer sheet.

The decisions may be difficult, but it is important that you complete all the items as thoughtfully as possible.

This test was developed by L.D. Mackay, Monash University, 1972. Copyright.

1. Penicillin was discovered by Alexander Fleming in 1928.

   A. Penicillin is an antibiotic.

   B. Penicillin is not an effective treatment for all bacterial diseases and a whole range of other compounds have been developed to treat bacterial diseases.

   C. Penicillin is extensively used by doctors in the treatment of bacterial infections.

   D. Pencillin is an effective drug because it can kill certain bacteria which are harmful to humans.

2. If a broad bean seedling in a pot is placed in a box, and light is only allowed to enter the box through one small hole, the growing region of the shoot inclines to the light and finally the leafy tip of the stem emerges through the hole.

   A. This is an example of phototropism.

   B. The explanation of this is that light has a retarding influence on growth, so the side of the stem facing the hole grows more slowly than the opposite side.

   C. Pot plants being grown inside near windows must be regularly turned so that they retain their normal shape and do not incline too much towards the light.

   D. Light is not the only stimulus to which plants can respond; for example, some plants respond to contact and to gravitational stimuli.

3. Scientists have observed that in towns in which traces of fluoride occur naturally in the drinking water, there is a lower incidence of dental caries than in similar towns whose water supply contains no fluoride.

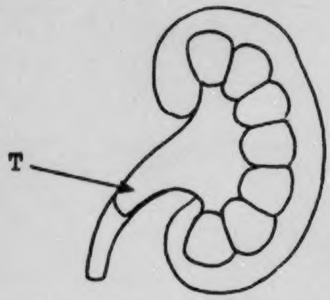
   A. The presence of fluoride in drinking water reduces dental decay.

   B. In some towns, fluoride is added to the water supply in controlled amounts.

   C. Fluoride in drinking water reduces the susceptibility of tooth enamel to decay which results from the production of acids in the mouth by bacterial action.

   D. Large quantities of fluoride in drinking water can be poisonous.

4. The diagram represents a section through a mammalian kidney.



A. Kidneys produce urine from waste materials extracted from the blood passing through them.

B. The kidney tubules are microscopic structures which extract materials from the blood and produce urine from them.

C. The tract T leaving the kidney is the ureter.

D. The extraction of materials from the blood by the kidney helps to maintain a steady state of water and salts within the body of the mammal.

5. The chemical reactions in a living cell are controlled by organic catalysts which are produced in the cell and are known as enzymes.

A. Enzymes are rendered inactive at high temperatures.

B. It is thought that in enzyme action the reacting molecules become temporarily attached to a suitably shaped surface of the enzyme, and are held in contact so that the rate of chemical reaction increases.

C. The action of enzymes is used to change starch to sugar in the malt house during the production of beer.

D. All enzymes are proteins.

6. The leaves of flowering plants have many small openings through which water vapour and other gases pass.

A. A continuous interchange of gases is essential for the processes of photosynthesis, respiration and transpiration in flowering plants.

B. The cells guarding the stomata are very sensitive to the presence of light and the size of the opening changes in response to changes in the incident light.

C. These openings are the stomata and occur chiefly on the undersurface of the leaves.

D. These openings open and close, thus regulating the passage of gases to and from the leaves.

7. Fungi are plants that are unable to make their own food and therefore must obtain food that has been made by some other living thing.

   A. Some fungi develop nooses by means of which they strangle nematode worms.

   B. Fungi do not contain chlorophyll and therefore cannot produce food from non-living inorganic material by photosynthesis.

   C. Fungi are food consumers rather than food producers.

   D. Yeast is a fungus which is used to produce alcohol from sugar in the manufacture of beer.

8. In order for fertilization to occur in flowering plants, pollen must be transferred from the stamens to the stigma.

   A. The pollen grains produced by the stamens contain the male cells while the stigma is attached to the ovule which contains the female cells.

   B. This transfer is called pollination.

   C. In most flowering plants self-fertilization is impossible.

   D. Growers frequently transfer pollen from one flower to another by hand to produce the desired variety of flowering plants.

9. The following table lists some of the features of each group of animals with backbones

ANIMALS WITH BACKBONES

FISH	AMPHIBIANS	REPTILES	BIRDS	MAMMALS
Body covered with scales. Possess fins and gills. No legs.	Skin damp and naked with no scales. Four legs.	Body covered with hard scales.	Body covered with feathers. Two wings, two legs.	Body covered with fur.

   A. The information in the table could be used to classify an animal with a backbone into one of five groups.

   B. Mammals are the only animals with backbones whose bodies are covered with fur.

   C. In addition to enabling the sorting of animals with backbones into groups, the table indicates possible evolutionary relationships between the groups.

   D. In the table, animals are grouped together according to their appearance and the way they live for ease of recognition and identification.

10. Rocks contain fossils which are the remains of organisms or parts of organisms that have been preserved, sometimes for millions of years.

   A. The majority of fossil remains are of aquatic organisms because they are more likely than terrestrial organisms to be covered by sediments.

   B. Fossil remains provide evidence of organisms that existed during different periods in the past.

   C. Geological evidence on the age of rocks in which fossils occur can be used to determine the order in which different organisms evolved.

   D. The fossilised remains of plants and animals are found embedded in rocks.

BIOLOGY PREFERENCE TEST

ANSWER SHEET

NAME ....
(FIRST NAME) (SURNAME - BLOCK LETTERS)

SCHOOL ....

FORM .... DATE OF TEST ....

IS THIS THE FIRST YEAR YOU HAVE ATTEMPTED LEAVING BIOLOGY? YES ☐ NO ☐
(Tick answer)

ITEM	MOST APPEALING CHOICE	LEAST APPEALING CHOICE
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

BIOLOGY ATTITUDE INVENTORY

Time allowed: 15 minutes

DO NOT OPEN THIS BOOKLET UNTIL YOU ARE TOLD TO DO SO

Fill in your name and other information on the Answer Sheet provided.

This is not a test or examination. There are no 'right' or 'wrong' answers.

What is wanted is your opinion, not what you think you ought to believe, or what other people want you to believe, but what you, yourself, do believe.

The answers to the statements should be given by circling the appropriate symbol on the Answer Sheet beside the number of the statement.

Draw a circle around

SA	if you STRONGLY AGREE with the statement.
A	if you AGREE with the statement.
D	if you DISAGREE with the statement.
SD	if you STRONGLY DISAGREE with the statement.
N	if you cannot decide whether you agree or disagree with the statement.

PRACTICE STATEMENT

41. Of all the school subjects I like Biology best.

If you disagree with the statement, you would circle D on the Answer Sheet.

If you change your mind about an answer, just cross it out and circle another one. Please give an answer to every statement.

This questionnaire is an adaptation by Arthur M. Lucas of a questionnaire originally developed as part of the Physics Evaluation Project, Faculty of Education, Monash University, 1967.

1. If a student was doing an experiment to investigate the effect of bright light on the behaviour of insects, the teacher should not tell him the results; rather the teacher should let the student find out for himself.

2. I enjoy reading books about biology.

3. Scientists are about as healthy and as physically fit as other people.

4. Scientists are rather eccentric persons who live in a world of their own.

5. Biology is worth doing whether you want to be a scientist or not.

6. A biology experiment is best when the instructions are all laid out in the textbook or laboratory manual.

7. The best way for students to learn about bacteria is to listen to a teacher, who knows a great deal about the subject, summarize the facts known about bacteria today.

8. Some laboratory manuals written for biology students to use in their experimental work don't give very definite instructions about what to do; instead they state some problem and expect students to work out their own procedure. This approach is quite good.

9. It is reasonable to ask students to learn laws like Mendel's Laws of Genetics, but they should not be told the laws before they do experiments with living organisms, they should try to discover the laws for themselves first.

10. I feel that it is a waste of time and effort for me to investigate any natural phenomenon for which explanations already exist.

11. Most experiments should be performed as demonstrations by the teacher, because the teacher is likely to get a better result than the student, and the student would therefore learn more.

12. Biology is just a dull grind for me.

13. Scientists are less warm and friendly than other people.

14. People with imagination would certainly be more successful as writers and artists than as scientists.

15. Hardly any scientists believe that God exists.

16. A biology experiment is good when students are given apparatus and told to solve a problem without instructions of how to do it. That's a real experiment.

17. I don't see why a biologist would want to spend most of his life doing experiments and analysing results.

18. I like the way biology challenges us to find out why things happen.

19. When you learn a new idea in biology it is very important to thoroughly learn the statement of it so that you know it word perfectly.

20. Once a law in biology is discovered it should not need changing.

21. I cannot see the point in studying topics like producers and decomposers in any greater detail.

22. For every research problem facing the practising biologist there is a well-defined procedure which allows him to obtain the answer.

23. Scientists are interested in religion to the same extent as other people.

24. I would enjoy having a job in a biology laboratory during one of my school holidays.

25. Men may have some of their most cherished ideas shattered by discoveries in biology.

26. Learning about the sort of thing that anyone can see happening makes biology boring for me.

27. Eventually all the laws of biology will be discovered and little further research will be necessary.

28. Some biology teachers occasionally give students problems to solve experimentally, even though the teachers do not know the answer to the problems themselves; this is quite a good idea.

29. The best biology teacher would be the one who tells you as many facts about biology as possible.

30. It is pretty well certain that biologists are reaching the limits of things a man can know.

31. Scientists are more devoted to their work than most people.

32. Scientists would not be interested in ordinary community activities.

33. In biology there are many problems for which biologists at present do not even know the procedure for obtaining the answers.

34. Scientists are less interested in art, music and literature than other people.

35. Scientists are peculiar people.

36. It is important for a biologist to be able to throw away widely accepted ideas and think without restriction.

37. The proportion of scientists that are happily married is about the same as in the rest of the population.

38. There is no end to the new things that a biologist may discover.

39. In the future, people will be surprised at how much biology has changed since now.

40. I find learning about bacteria most enjoyable.

BIOLOGY ATTITUDE INVENTORY

ANSWER SHEET

NAME ______________________________ ______________________________
(FIRST NAME) (SURNAME - BLOCK LETTERS)

SCHOOL ______________________________

FORM ____________________ DATE OF TEST ____________________

IS THIS THE FIRST YEAR YOU HAVE ATTEMPTED LEAVING BIOLOGY? YES ☐ NO ☐
(Tick answer)

	PRACTICE STATEMENT
41	SA A [✓ D] SD N

1	SA	A	D	SD	N	21	SA	A	D	SD	N
2	SA	A	D	SD	N	22	SA	A	D	SD	N
3	SA	A	D	SD	N	23	SA	A	D	SD	N
4	SA	A	D	SD	N	24	SA	A	D	SD	N
5	SA	A	D	SD	N	25	SA	A	D	SD	N
6	SA	A	D	SD	N	26	SA	A	D	SD	N
7	SA	A	D	SD	N	27	SA	A	D	SD	N
8	SA	A	D	SD	N	28	SA	A	D	SD	N
9	SA	A	D	SD	N	29	SA	A	D	SD	N
10	SA	A	D	SD	N	30	SA	A	D	SD	N
11	SA	A	D	SD	N	31	SA	A	D	SD	N
12	SA	A	D	SD	N	32	SA	A	D	SD	N
13	SA	A	D	SD	N	33	SA	A	D	SD	N
14	SA	A	D	SD	N	34	SA	A	D	SD	N
15	SA	A	D	SD	N	35	SA	A	D	SD	N
16	SA	A	D	SD	N	36	SA	A	D	SD	N
17	SA	A	D	SD	N	37	SA	A	D	SD	N
18	SA	A	D	SD	N	38	SA	A	D	SD	N
19	SA	A	D	SD	N	39	SA	A	D	SD	N
20	SA	A	D	SD	N	40	SA	A	D	SD	N

APPENDIX II

A CATEGORY SYSTEM FOR THE ANALYSIS OF TEACHING STYLES (CATS)

A CATEGORY SYSTEM FOR THE ANALYSIS OF TEACHING STYLES
(CATS)

1. DEFINITIONS OF CATEGORIES

FIRST LEVEL CATEGORIES

TEACHER INITIATION (Using initiatory behaviour to begin a teaching cycle or to extend a previous teaching cycle)

Presentation includes all instances where new information, substantive or managerial, is given in a way that is not immediately concerned with the various possibilities of response. The information can be fact or opinion, and may deal with knowledge, skills or attitudes. However, if information is given as the result of a request, this should be considered as development.

Questioning is another initiatory move and includes behaviours which are intended to elicit an answer. Some questions ask for information about content and procedure, and require the use of memory alone, others call for the use of higher-order intellectual skills (Bloom, 1956). Many questions asked in a classroom have a known 'right' answer, however 'open' questions do not. Most solicitations requiring verbal or written response are included in this category, but rhetorical questions are not. Some imperatives such as "Tell me how many legs an insect has" are regarded as de facto questions.

Direction includes behaviours which are intended to structure the behavioural response, either in the expectation of compliance as when a command is given, or by providing alternative courses of action as when a suggestion is made. Most solicitations requiring non-verbal response are directions. Requests which typically begin with the formal interrogative "Would you please ...?" are regarded as de facto directions.

TEACHER RESPONSE (Using response behaviour to close the teaching cycle)

Development is a response behaviour and includes those acts which serve to modify what has gone before, by clarifying, interpreting, summarizing, expanding, or synthesizing. Development takes place in the context of what has gone before, but need not necessarily be directly elicited by the preceding event. Evaluation of, or opinion about, substantive content and process is included, but not the evaluation of pupil responses.

Evaluation of pupil behaviour includes those acts which could be classified as 'evaluation' in Bloom's Taxonomy, as well as instances of positive and negative reinforcement. Judgements of the correctness or appropriateness of a pupil's behaviour are sometimes made without any discernible affective connotation, usually in terms of public or external criteria (such as accepted definitions, conventions, or empirical verification): these could be regarded as cognitive evaluation. Where the judgements have discernible overtones of personal criteria, and could be described as 'encouragement, praise or approval' or 'criticism, reproof or disapproval', the evaluation is affective.

TEACHER NON-VERBAL BEHAVIOUR

Where they occur together, non-verbal behaviour is used merely to elucidate the intent of the verbal behaviour. Most non-verbal behaviour is congruent with the accompanying verbal behaviour: some, such as sarcasm, is not. All other teacher non-verbal behaviour that appears to have an instructional function should be included, including the use of audio-visual media to communicate with pupils.

PUPIL RESPONSE

Included in this category are all verbal responses by pupils to some initiative or action taken by the teacher (such as the asking of a question or the giving of a direction).

PUPIL INITIATION

Included in this category are all pupil verbal behaviours not directly stimulated by the teacher. Unprompted comments, observations, and questions asked by pupils are all included.

NON-CATEGORIZABLE

This is a residual category to contain all behaviours, verbal and non-verbal, that have no obvious instructional function, or which for some other reason can not be categorized.

SECOND LEVEL CATEGORIES

OPEN AND CLOSED BEHAVIOUR

Open behaviour on the part of a teacher is designed to lead a pupil to find out things for himself rather than to accept another's word. It accepts the pupil's frame of reference, and would be expected to promote pupil-initiative and independence of thought and action. Pupils are encouraged to question, to develop critical capacities, skills in inquiry processes and to be creative. Closed behaviour is 'conservative', is authority-centred and designed to promote uncritical acceptance of the present state of knowledge, methods, and procedures. Pupils are encouraged to adopt the dominant values and outlooks of adult society as interpreted by the teacher.

POSITIVE AND NEGATIVE AFFECTIVITY

These categories include the commonly used categories of encouragement, praise or approval on the one hand, and criticism, reproof or disapproval on the other. These behaviours can be verbal or non-verbal and will tend to reinforce the pattern of open or closed behaviour with which it occurs. Other evidence of good-humour such as smiling, laughing and joking, as well as evidences of ill-humour such as scowling, frowning and curt dismissal should also be included.

LEVELS OF QUESTIONS

A memory question is one which requires only the recall or recognition of material already learned. A higher-order question is one which requires the processes of Comprehension, Application, Analysis, or cognitive Evaluation (Bloom, 1956) towards a known answer; a convergent thinking, or convergent evaluative question. An open question may have no known answer: it may require the expression of an attitude, the proposal of a plan, some creative, divergent, or unique response.

THIRD LEVEL SUBSCRIPT

TEACHING GROUP SIZE

Teachers may interact with the class as a whole, as when teaching from the front of the class, or alternatively with an individual or a small group of from two to four pupils.

FOURTH LEVEL CATEGORIES

MANAGERIAL AND SUBSTANTIVE BEHAVIOUR

Managerial behaviours are those which are primarily concerned with the organization of the instructional environment, such as routine organization of the classroom and laboratory, control of student behaviour and the building of a congenial classroom climate. All acts of positive and negative reinforcement should be included. Substantive behaviours are those which are primarily concerned with the presentation and development of subject matter and the acquisition of related skills and attitudes. These statements and actions are based on efforts to achieve the basic teaching objectives, and may be as much concerned with the procedures of inquiry (process) as with the products of inquiry (content).

---

2. RECORDING

The CATS may be used to code classroom behaviour from direct observation or from annotated audiotape or videotape recordings. However, the procedures described here have been developed for a research purpose, and are designed for use with annotated transcripts and audiotape recordings.

1. The radio-microphone should be fitted to the teacher before he/she has any contact with the class for the lesson. If fixed microphones are used, these should be in position, and tested, before the start of the lesson.

2. The observer should sit where he/she can clearly see the classroom interaction, particularly the teacher's behaviour, and yet be as unobtrusive as possible. This usually means sitting at the back of the classroom. The observer should wear earphones to monitor what is being recorded on tape and to prevent him/her from being influenced by audible clues that may not be picked up by the microphone(s).

3. The observer should record all non-verbal behaviour that would modify the intent of the teacher's verbal behaviour (e.g., sarcasm, approval, disapproval), and make a written description of the teacher's behaviour during periods of silence (e.g., handing out apparatus, supervising written work, using audio-visual materials, assisting with laboratory work, out of classroom). The observer should describe, and note the time of, all instructionally-significant non-verbal behaviour that occurs before the first word is spoken. The time of the first spoken word should be noted so that the written notes of the observer can be synchronised with the audiotape record. Similarly, the time of other important changes in class activity should be noted (e.g., film used, students moving to laboratory benches, announcement over the public address system, written test administered).

4. Written transcripts of the audiotape record are then made. The transcript can be divided up into five-second segments and annotated with the observer's written record of non-verbal behaviour. The transcript can be divided into segments by replaying the audiotape together with a tape which gives an audible signal every five seconds, and marking the transcript where the signal occurs. If the signal does not occur at the end of a phrase or sentence this mark is made at the nearest logical break in the statement being made. For additional convenience in locating episodes in the transcript the signal tape can have a signal of a different pitch every minute.

3. CODING

Each five-second segment of behaviour, verbal and non-verbal (or each statement, if there is more than one statement in the interval), is then categorized according to its inferred intent or instructive purpose. The following decisions are required:

FIRST LEVEL DECISIONS

To which of the categories 1 to 8 does the behaviour belong? If the behaviour can not be categorized, record 1 in column X.

SECOND LEVEL DECISIONS

If non-verbal, is the behaviour open, closed, positive, negative, or audio-visual in nature? Record 0, 1, +, -, or 'a' in column 1.

If development or presentation, is the behaviour open or closed? Record 0, or 1 in column 2 or 4 respectively.

If evaluation, is the evaluation positive, negative, or neutral? Record +, -, or 1 in column 3.

If questioning, is the question memory, higher-order, or open? Record 1, h, or 0 in column 5.

If direction, is the direction open, closed, positive, or negative? Record 0, 1, +, or - in column 6.

If pupil behaviour, is the behaviour open, closed, positive, or negative? Record 0, 1, +, or - in column 7 or 8 as appropriate.

THIRD LEVEL DECISION

Is the teacher interacting with one to four pupils? If so, subscript the entry with 'i'.

FOURTH LEVEL DECISION

Is the behaviour managerial in function, or concerned with the development of substantive content or skills? Record M, or C in column 10.

GROUND RULES AND ADDITIONAL NOTES

All behaviour should be coded in its context. For example, the answer given, and the teacher's reaction to the answer, are important in deciding the intended level of a question. The form of a question may appear open, but the teacher's response to the student's answer may clearly indicate that the teacher was looking for one particular answer. On the other hand, a teacher may ask a student to explain his previous answer to stimulate a more creative response.

1. Opening verbal formalities such as "Right", "Er-r", "Now then!", are interpreted as having a function similar to "I am about to speak, please pay attention": code 61M.

2. Dictation of notes: code column 2 if it is a summary of work done, code column 4 if it is the presentation of new material.

3. Repetition of a student's answer, or the comment "OK", or "Right (meaning "That's right")" following a student's answer: code 31_iM.

4. "Sorry", the equivalent of "Yes you are right (and I am wrong)": code 31_iM.

5. Silent assistance to individual with work: code 11_iC.

6. Walking around the room silently while students are working is regarded as supervision: code 11M.

7. Pause waiting for a student's response: code 11_iM.

Appendix II. Encoding sheet for CATS

1	2	3	4	5	6	7	8	X	10	1	2	3	4	5	6	7	8	X	10	1	2	3	4	5	6	7	8	X	10

J.H.T. 1974.

APPENDIX III

ANNOTATED TRANSCRIPT OF SAMPLE LESSON
(LESSON 21)

The end of each five-second interval is marked by an oblique stroke (/), and the end of each elapsed minute is marked by a number in brackets, e.g., (24). Annotations are made in brackets.

ONE-SIXTH SAMPLE CODING OF LESSON 21

Note that the five-second intervals which are coded here are not successive intervals, but represent each sixth five-second interval throughout the lesson.

ANNOTATED TRANSCRIPT OF LESSON 21

Lesson Topic: Web of Life (1969 edition) Exercise 5.4
Diversity in Plants

Purposes:
1. to become familiar with the diversity of plant structure found in nature,
2. to become aware of the features used to classify plants, and
3. to gain some knowledge of the major groups commonly used in plant classification.

(Student's Manual, Part 1, A.A.S. 1969)

T. Janice, shut the door please. (Waiting while door is being closed) ////

T. Now could we have your attention everyone please, / all facing the front. The exercise we're going to do / this morning is Diversity in Plants. / (Teacher writes on blackboard, "Diversity in Plants".) /

T. Now you may or may not know the meaning of the word 'diversity', / if you don't there's a dictionary in the cupboard and probably by the end of the morning / having done the exercise you'll know what it's all about.

Around the room you'll notice / that there are plants and various bits and pieces scattered. They're numbered from 1 starting here, (pointing) (1) to 2, 3, 4 right round the room. Now you will work in your pairs, / and you'll examine these different sections, / plants at the various spots. In your data book I want you to put this heading / (Teacher writes on blackboard "Purpose".) and underneath the purpose of the exercise. / Now I'm not going to tell you what the purpose is, at the end of the exercise you're going to insert your own purpose / because by that time you should know exactly why you've done in this exercise. / Then each person will take to this section where they're going / a data book and numbers inserted in the data book / from 1-14 with a space of about three lines between. / Because there are 14 specimens to have a look at, you have to say something about each one of them. / Now just leave your data books for a moment please until we've finished talking. / (To a student, insistently) Would you put that away! (2)

T. Well, a little while ago / we had a look at a whole series of animals, now John / what sort of things did we observe about that whole collection of animals? /

S. Structural differences. /

T. Structural differences. O.K. / Mary. Can you give us an example of a structural feature or a structural difference between two types of animals? //

S. You mean the bones of the two fishes? Do you want that? /

T. Yes. /

S. One's bony and the other's cartilaginous.

T. Could you actually see whether a fish were bony or cartilaginous? // You're more concerned with the gill-slits. All right? (3)
Now Sue, do you think we could use structural features in trying to sort out our plants into groups? / What sort of features might we use?

S. The shape of the leaves. /

T. Mm! Mm! (Encouragingly)

S. Colours.

T. Mm! Mm! (Encouragingly)

S. If they've got flowers. /

T. Good. Well that'll do to give you an idea. Fine. Now as you move around / these spots you have to write in under No. 1, for example, / all the features about this particular specimen which you think will help you put it into a particular group. / Now on your desk in front of you between two people there should be the old manuals. / We'll turn up Exercise 5.4, page 4. //// Exercise 5.4. Page 4 and 5. (4)
Now having decided / and written down some of the important structural features of your plant, / you then use this key and try and work out to which big group your plant belongs. / Now these are just ordinary, every-day sort of names, some of them you will have heard before, some of them you will have not. / So just work carefully through this structure and see if you can find / which group of plants No. 1 belongs to and you write it in there. / And you proceed round the room. Now, because there are 14 spots to go to and about 11 pairs of people, / we shouldn't have any congestion at all once we get moving. / Now, even though you have seen some of these plants before and you think you know all about them, / I want you to look at them carefully and notice as many structural differences as you can, or important structural features, / before you start to use your key and identify them. This'll probably not take you the whole time. / If you have finished, I want you to come to me and get a sheet of questions (5) which you have to answer further work that has to be done. You'll

notice / round about the room that there are books scattered, please don't use them at the moment, they're for use when we start on the questions that are on our sheet. /

There are lenses, hand lenses about, to use, / stereo microscopes to use, and if you want any other equipment / that's going to help you with your observations, just let me know. Now, has anyone got any questions before we get moving? /

T. All right, off you go. (Students move to laboratory benches. Teacher moves around class talking with small groups as they work.) ////// (6) //

T. Now where are you folk going to? /

S. (Inaudible, but probably with the sense of "To find a place to work")

T. Anywhere where there's a gap. Just fit in. ////

Now let's spread these plants out so that we can get a good look at them. / Notice anything about their groups to start with? / Anything rather obvious?

S. Colour.

T. Colour, good. /

S. Thickness of their leaves.

T. Mm! Mm! (Encouragingly)

S. Some are thicker.

T. Uh-huh! (Encouragingly) /

What sort of features might you be going to look for? (7) What's that there?

What's that?

S. A root. /

T. Is it really a root? What's your definition of a root. / What is a root? You have to think and decide whether that's really a root, or not. /

T. What do you think it is Jane?

S. Type of flower? /

T. Well, is the black part growing out of that? /

S. Yes. /

T. Mm mm! So that is stuck hard on something.

Now this is a seaweed that's been washed by the oceans. / What might the function of that thing be, do you think?

Gary? /

S. I don't know.

T. What? Imagine your seaweed in the ocean flapping away.

S. Well, you'll need to keep it in the water. /

T. Good. It's called a hold-fast, attaches to the water. / No roots.

S. (Inaudible)

T. No roots, just a holdfast. It's called a holdfast. Right? /
Well, anything else about this? What other sort of questions would you ask yourself? (8) (Interference on tape. Student's comment inaudible) We're looking for structural features / which might help you group this in a certain group of plants. / We've decided no roots, right. Anything else? /
Gary, anything else? /

S. It's got long strandy parts on it on this black leaf. /

T. Are these what you consider a real leaf, a typical leaf? /

S. No, not really.

T. So, no roots, no leaves. / What else?

S. No stem. /

T. There is a stem there, a type of stem. What else is often found in flowers-/ I've told you now - in plants?

S. Flowers.

T. Flowers. No flowers? / No.
Well, write down all the things it hasn't got as well as things it has got and then look at your key / and see if you can find out what it is.
(9)

T. What have you got here Stephen? /

S. (Inaudible) ... algae.

T. What's it look like?

S. It's thick strands. / See the cells.

T. Mm! How are the strands and cells related? /

S. They're just ... eh?

T. How are they made up?

S. Oh, the single cells make up strands. /

T. Good.

S. Yes, we worked out what it was. We didn't know what it was. / (Teacher moves on to another student)

S. (Inaudible comment)

T. Now, what have you got here? / Margaret.

S. Kind of bacteria.

T. What makes you say it's bacteria? / Is this from your key, is it?

S. Yes.

T. Mm mm! Have a look down there down the microscope and tell me what you see. / Just in your own words. Or draw it on your book if you like / and show me it on your book./ They're all a bit sandy from the seaweed. (Referring to other plant specimens)/ Do you see any that look sort of like this - big one and little one. (10) You have a look, look carefully through the field. // No? (Teacher looks down microscope) ///// Now you look right down the middle of the field / and you can see / Let me show you with your pencil. / There's one just there just like that / (Drawing) and there's another big one just down there. / Looks as if that one's coming off. (11)

S. (Inaudible question)

T. You'll find that out when you get your sheet in the next section, / when you get .... You don't need to know, no need to know what it is, / just look at the plant structures. / Now, do you see what I mean? / Well? Do you think that little one is being produced by the big one as a means of reproduction? / What sort of a word might you give to that type of reproduction? / Is the thing dividing equally in two? /

S. No not quite.

T. Well what would you call that little thing that's coming off? /

S. I think it's ....(Inaudible answer).

T. Yes, but something in the plant world, / something small, that grows from something. /

S. Bud.

T. Bud. Good. That's right, it's budding. Reproducing by budding. Fine. / (Moving on to another group)

T. Well, what have we got here? (12)

S. They're two different there, they're both ..../

T. Yes, they're different, but they're in the same big group, this one has got these little structures. / Later on, in the autumn, this one will grow them too.

S. Oh. /

T. Now, have you noticed these little structures, they're pretty important. Have a look at them with your lens, how would you describe them Louise? /

S. They're very thin. /

T. And how many parts?

S. There's a few. /

S. They've gone together. There's one in the bud. /

T. Yes, the individual structure itself, how many parts would it have? What's on the tip? /

S. A little trunk ....

T. Mm! A little knob on the tip and a long filament isn't it? / And they're a pretty important part of the plant. /

S. What's that on there?

T. That's just a bit of rubbish.
Now has this thing got any roots? /

S. Uh!

T. Would you like to have a look? (To one of the other students in the group)

T. Take a little bit (13) and just tear it apart, see if you can see any roots. Perhaps the dry one you might manage better with. /

S. Yes, it has ...

T. Yes, it's got short roots. Any leaves? /

S. No. Some furry stuff.

T. Have you looked at them with the hand lens? / What's that fur made up of? Furry stuff as you call it. /

S. Oh, I suppose leaves.

T. Tiny little leaves?

S. Yes, very small.

T. Mm mm! / Any flowers?

S. No.

T. Right, well, write down what it has got, / what it hasn't got and see if you can identify it. / (Moves on to another group)

T. Now this is an easy one, isn't it? /

S. Yes.

T. What's it got? /
(Interference on tape)

S. ...instead of coming up ... (More interference) .... //

T. What else might you expect (14) it's got that you can't really see.

S. Roots. /

T. (More interference on tape) Well, would you be happy to say it's got roots or .... /

S. You hold it ... (Inaudible conversation between students and teacher due to faulty tape) / ....

S. Do you want it?

T. Well just dig around the bottom with your pencil. /

S. (Inaudible question) /

T. Good. Well all right write that down, that's an important thing. / (Moves on to another group)

S. What would you call those little brown lumps in this? /

S. What would you say they were? /

T. Well supposing you have a try, what would you describe them as? Do they look like little cups / or bowls? All right?

S. Yes.

T. Little cup-like structures. Now what do you think they might be used for? /

S. Seed pods?

T. Yes, they might be seeds. But a general term that involves seeds? / Actually these don't produce seeds. But what's a general term that we could use that involves seeds? (15)

S. Reproduction.

T. They might be reproductive parts, mightn't they? / You don't know but you can write this down.
What else have you said about it Fleur? /

S. Oh, parasitic. It's got parasites .... /

T. Mm! / What else?

S. (Inaudible comment) /

T. Mm! Any roots? /

S. No. /

T. What other structures might you look for? What are the other common features in plants? /

S. No branches.

T. Mm! No branches, no roots. What else? Any leaves? /

S. We can't see really. /

T. Well, you got to decide what your definition of a leaf is, / it depends on your definition of a leaf.

T. Oh good, right, no leaves. (16) So you can go on. / (Moves to another group)

T. Now, what about this one? / What features has it got and then we'll look at the ones it hasn't got. /

S. That the leaves are in the form of long prickly needle-like things.

T. Good.

S. And they're all in bunches. /

T. Good. Mm mm!

Now, here's a bunch of leaves with rather a special sort of structure on these. / Both come from the same plant. Now that's a young one of that isn't it. / Now what might these be do you think?

S. Very young ones.

T. No. / What are these involved with, what part of the life history of the plant? /

S. I've got the seed ....

T. Yes, so what's the general word we use?

S. Reproduction. /

T. Good, yes - reproduction. Do you think this might be anything to do with it? /

S. Which one? (Inaudible comment) ... reproduction.

T. In reproduction / what, in the high animals and plants, do you need if reproduction's going to take place? (17)

Male and female.

T. Well, got any ideas about this one? /

S. That one?

T. You have a think.

S. That's a male one.

T. Could be. / Well, fine you've got two reproductive structures on your one plant. / (Moves to another group)

T. Yes Janice?

S. How would you do this about the mildew, it says .... /

T. Let me come and have a look at it. /

Gary, you tell us exactly what you see. /

S. Just seems to be two little things on the end of it.

T. Mm! Then focus up and down, / what is beneath these things that have got the little knobs on them? You see a network of anything? /

S. Oh, there are fibres. /

T. Yes. Right, So you see a mass of fibres or threads if you like to call them / and then these little ones that come upwards with a knob on the end. / Now, any roots in this do you think?

S. Yea, there would be roots. (18)

T. Well, you've got to see them haven't you?

S. Oh, that'd be like the other ones down below, wouldn't it? /

T. Well, make a note in your book if you think those masses of threads underneath are roots, you put "possibly roots". You don't really know, do you? /// (Moves on to another group of students)

T. What do you think the little knobs on the ends of the aerial structures, / the ones that go upwards, might be involved in? /

S. ... sort of one tiny little thread. / Might be a bud?

T. Well what are flowers usually associated with? What process in a plant?

S. Pollination. /

T. Yes, or?

S. Reproduction.

T. Good, reproduction. / Do you think this might be involved with reproduction at all, those little black knobs on the end of the stalks? / Might be? So when you answer your question/ sheet you'll be able to find out a bit more about this, but I want you to notice the two main structures, (19) what would you say the two main structures on that board were? /

S. Oh, the stems and the leaves.

T. Yes. All right. / That's all you want to know. There are no flowers, no roots, no leaves. / Hasn't got much has it? / (Comes up to another group of students who are talking)

S. She'll swallow that, won't she! /

(Laughter)

T. And what would you say about these? /

S. The leaves are wide.

T. Are they leaves? /

S. Oh, well.

T. You have to be very careful of what your definition of a leaf, don't you? Are they really leaves? /

S. No. Not really.

(Inaudible comment from another pupil) /

T. What was that Lana?

S. Do they tear? /

T. You can tear it, yes.

S. They stretch don't they?

T. Are they leathery?

S. Yea.

T. Rubbery?

S. Rubbery. /

T. Mmmm! (Encouragingly)

Some of them are. What about this one?

S. Oh, that's thread-like. (20)

T. Mm!

S. No, that doesn't stretch.

T. Mm! What do you notice between this one and that one? /

S. Well, they're different colours.

T. Mm! Good. /

S. The one there has ... one has these ....

S. What would you call this? /

S. Bladders.

T. They're little bladders, call them bladders, and some of these they use for reproduction and some of them, they just float. /
Now what about this thing here, / there's a bit of shell there. Do you see the shell?

S. Oh, that looks like a mussel / doesn't it?

T. Mm! That's an empty shell, fine, now look at the seaweed / connection. What is the relationship between the shell and the seaweed? /

S. Well, they've got, er, type of seed pods.

S. They're stuck on to it. /

T. Well, yes, but why do you think the seaweed has latched itself onto these. / Well they're not really like weeds but they're branch processes because they stop where the shell stops. Don't they? / Why do you think the seaweed has done this? (21) Would it be an advantage to the seaweed if, with the pounding of the sea, it was continually moving around? /

S. That would ... I suppose. / I don't really know. /

T. If. Usually the shell would be attached to what? /

S. The rock.

T. Rock.

T. And the seaweed's attached to the shell, so the seaweed's going to move far?

S. No. / It stays in one place.

T. Yes, it holds it and it's called a holdfast. Right? Would you call these real roots? /

S. No, they're just stuck on sort of things.

T. And what about this stem, / do you think that's a real stem? You've got to decide. /

S. Oh, I suppose so.

T. It's a short stem isn't it. / Any flowers or other structures?

S. No. It hasn't got any. /

T. Mm! Well, look for the things it has got and the things it hasn't got, mm! (Moves to another group) / (22) /

T. Now have you people been to all the other spots? /

S. The book shouldn't be there. I've been to no one, the bread one. (Where mould was growing) / We're going to do that again.

T. They've nearly finished up there anyway, I think.

(Inaudible comments from some students)/

T. Have you used your key to identify some of these? / Any troubles with any of them?

S. No. 1., the bread one. / Especially the mould.

T. Now what did you find out? Plants with roots, mm!

S. No. /

T. Green in colour?

S. Yes.

S. Or greenish.

S. It's greeny blue. /

S. No, there's greens and yellows.

S. Maybe you weren't sure about the colours, or something.

S. Little black dots.

T. Mm! / Then where did you go?

S. We didn't ....//

T. Not green (23) or greenish in colour. Individual parts microscopic?

S. Yes. /

T. Cells minute even under high power. Did you look at this under the high power? /

S. Yes. We did that under high power.

T. This under high power, the stereo? /

S. I'm not sure about that

S. No, we didn't have it on that.

T. What was your magnification for upwards when you're using high power / Fiona?

S. Mm!

T. What objective and what ocular would you have? /

S. Twenty-forty.

T. What ocular? What is the highest ocular, Jan? /

S. Ten and forty.

T. And give you how many magnifications? /

S. 400.

T. 400. Now do you get 400 magnifications with your stereo Mary? /

S. No, I don't think so.

T. Nowhere near. / So it's nothing as tiny as that. / (Teacher moves on)

S. Here.

T. Fungi and yeast or ...? /

S. Yeast.

T. Is it a unicellular animal? (24)

S. No!

T. No? So you're stuck about there. That's as far as you can get / so you really haven't been able to ....

S. Yellowish-brown marks ....

T. That's a whitish colour isn't it? / Except for the black knobs.

S. Yes.

T. So, you make a note that you're stuck at that part of the key. / So later on we'll discuss this one. / There you are. (Moves on)

T. Now, have you done your liverworts along the road here? /

S. Well, we're not going great because there was someone there. We've got to go back. /

T. I see. /

T. Why is that apart? /

S. Because it's ....

T. Eh? /

T. What are the common structures of plants / that this one does not have, Jackie? What are some of the common structures / of plants that this particular one doesn't have?

S. Well, it doesn't have peepholes, it looks like. (25)

T. Mm mm! What else? No roots, no leaves. /

S. No flowers.

T. No flowers.

S. No seeds. /

T. No seeds. Right. (Laughter)
And yet it's quite a definite sort of plant isn't it? / What about these little things here?

S. Those brown things? /

T. Little cup-like structures, what do you think they might be? / Quite a few of them, especially on the one that Lynne's got. // They're quite a distinct feature aren't they on the thing, / Think they might serve any purpose Lynne?

(Comments from pupils inaudible because of interference on tape) /

T. Well how do you think this particular plant / how do you think this thing reproduces itself? /

S. (Inaudible answer)

T. Mm! Possibly by those small structures, if you look at them, under the microscope. (26)
That section. Look at them carefully under the microscope. You see just the reproductive structures / in those little cup-shaped things. //

T. Now, just check to see if the slides still O.K. / Yes. Here's the same thing microscopically. / This is a bit of it under the microscope. / Now let's have a look at this one / just to check that the field's all right. / Now will you focus carefully with your fine adjustment.

S. I've got ....

T. Oh! / You'll have to focus carefully with your fine adjustment. / What is the predominant organism in there? Have a look and tell me. / Here's your fine adjustment. (27) /

S. (Inaudible question)

T. Little round green ones, / masses of them. Now they came from this bit of lemon.

S. Oh, and they're all grouped in one shape too. /

T. Mm! So have they got any roots or limbs, / you know stems? /

S. (Inaudible) ... type of stems, yes.

T. No. Well what would you say that plant was? / How could you describe it? How many cells do you think each one's made up of? /

S. (Inaudible request from pupil) /

T. Just a minute Jan, I'll be with you in a moment.

S. (Inaudible) ... the blackboard. /

T. What's a word that we could use that means the plant's made up of one cell?

S. Proto - Protoz ... Oh, no! /

T. Those are animals, yes?

S. Yes.

T. What's a prefix we sometimes use that means one? /

T. Ever heard of uni? (28)

S. Yes.

T. Unicellular. Mm! That's what a unicell looks like there. / (Moves on)

S. Is that a flower?

T. Well, in your book, does that look like a flower? /

S. No.

S. Yes.

S. No. / Well except for that little ... They could be petals. /

T. Right. Well, what is a flower involved in, in the process of a plant?

S. Reproduction. /

T. Mm! You think that this might be a reproductive structure of this particular thing?

S. Yes, the seeds are falling out of it.

S. Yes. /

T. Mm mm! Actually we don't call them seeds, you'll learn a little bit later what they are. We call them spores, but have you looked at them under the stereo microscope? / Did you see along these gills, as they're called .... little specks .... (interference on tape) /

S. Little drops of water.

T. Mm! If you look carefully you'll see those. / Any roots or ...?

S. (Inaudible comment)

T. Where's the roots? /

S. (Pointing) O.K.? /

T. Now is that your idea of a root? What's your definition of a root Jan? (29)

S. Oh! / Well, it's the long thing you know that goes in the soil, / it's what the food goes through.

T. Mm! What other use, Gary, does a root / have besides taking in water and certain minerals from the soil for your plant? /

S. Keep the plant steady in the ground.

T. Mm! It anchors it. Now, has this got a root system / that you can see? It's got a few little fine threads / but is this your idea of a root system?

S. It looks like it's starting to grow. /

T. Well have a look. These are about to die, so they've done all the growing they're going to do. / A few of those little hair-like things on that one.

S. Could have been roots that are starting to die off.

S. There's something in there! /

T. That's a bit of coloured stone I think. /

S. Could it have been there when the root started to die off? /

T. I think if it had a rooting system, this one has definitely parted company (30) from whatever it was. This one hasn't. This one has still / got its little fibres sticking out. I don't think we can really say this has got a root system. / So just write down what you can see, and when we come to have a look at our books / we'll look at the other organisms and see if we can find out what it is.

T. Some of these you may not be able / to clear up on your key. Well if you get stuck, you put a note in your book as to how far in the key you got to / and why you got stuck. /

T. Now as soon as you've finished your observations you can get one of the question sheets. ///// (Teacher moves away from group)

S. Please miss, do I do this one?

T. No, that one's done. (31)

S. We're meant to have the green stuff? /

T. Here. (Interference on tape) ///

S. I didn't realise we'd got .../ (Teacher moving on to another group)/////

T. Now, what have you discovered about this one? /

S. Well, it's got broad leaves. (32)

T. Are those really leaves? / What's your definition of a leaf? / Think they're really a leaf? /

S. (Inaudible) ... a bit, unless they're flowers. /

T. Well how would you describe a leaf in an ordinary ... /

S. Something that comes from the stem.

T. Mm mm! Short stem perhaps. / Some of them come directly from the branch without a stem but what else is usually in a leaf, structurally? /

S. Veins.

T. Mm! Any veins in these? /

S. The stem's not (inaudible) ... (Another student talking in the distance).

T. Are they attached to a stem?

S. No! They grow directly out of the ground. /

T. Well, all right. Are you going to report a true leaf, / or a leaf-like structure or what?

S. A leaf-like structure. /

T. What about on them, notice anything growing on them?

S. Yes. Hairs.

T. Hairs, yes. (33) Lend me a pencil a minute, what about these long things? (Pointing) /

S. They hold the stems.

T. What about these? /

S. Actually, it looks like a new shoot coming out of it.

T. Mm mm! Right. / Any flowers?

S. Well not really. /

T. Any roots?

S. Well, I guess they must have, to grow! /

T. You sure? What's your definition of a root? / Which side has got any roots? (Long pause while student examines specimen) /////(34)/////////

T. What can you see in this? /

S. Stem, branch, they're branches.

T. Well, hold this down low, / a bit deeper into the culture, and see if you can see anything below that branching structure. (35) / What do you think's actually right next to the bread, what's growing right next to the bread? // Put this one under. / Now can you see where the knife is? /

S. I can see little white things.

T. What are they? /

S. Little branch things, threads.

T. Threads, filaments? / Lot of them? /

S. Um! No, not really.

T. Focus up and down a bit. /

S. (Inaudible comment) /

T. What would you say are the two main structures you can actually see in this particular plant? /

S. Um! These little black things. (36) There's some white ones too and they're .... /

T. Mm! Good, any flowers. Any true stems? /

S. No.

T. Any colour other than the black. /

S. Um! White.

T. White. Mm mm! There's no green though is there in these, no? / Make a note of all the things that aren't there.

S. (Inaudible question) //

T. See, you can take this apart if you want to.

S. What's the different colours in places, / you know different shades all of the same two colours ..../

T. Well, if we move the piece of bread around, (Teacher moving bread under the microscope), you'll get a black patch underneath, or a green patch. / These are all different moulds mostly growing like this. This is one sort. It's very uncommon bread; / these are other types of moulds. Some of them are bacteria; some of them moulds. / Bread makes a very good culture medium, here's a green one over here. (37) You can see the young growth is white, the dark growth is green. Now that organism is very similar / to the one that's down on the lemon, have you had a look at the lemon down the corner? Mm mm!

S. What's that dark one? /

T. That is just an older one of this; it's just covered in black spores. / Come over here, you can't see. / Like a great field: it's like aerial structures with the / black spore case on top.

S. Mm!

T. You've got an old patch there you see. / There's a young patch, and that's the old patch. /

S. Now that patch is green. /

T. Mm! See how it's spreading right over the bread. Here it's spreading, starting to spore and spreading right over the bread. /

S. What happens after it's spread right over the bread?

T. Well, what do you think? / How do you think this thing gets its food?

S. From the bread itself. /

T. Yes. Well, given long enough, what will happen to the bread? (38)

S. It'll just go.

T. Yes, it will all disappear. /

S. Will it die?

T. Well, it will die, but inside these little black cases are spores, / we'll learn a lot more about spores later, which are very resistant, and this new mould will grow from these spores in the spore-case. / Out of this. (Pointing)

T. Haven't you seen this one / on a piece of mouldy bread at home at any time?

S. Yes.

T. It's rather an interesting plant, / with none of the usual structures that plants have. You might say "Why call it a plant / when it doesn't seem to have any of the ordinary sort of structures we think of that are on plants?" /

S. (Inaudible question)

T. Mm mm! /

(Observer changes sides of audiotape cassette) x/x/x/

T. There are a few features about them, but you can refer back to them if you like. (39)

S. Shall we go back (Inaudible) ....? /

T. No, you can start on your sheets. /

S. Oh!

S. You know that other one, when we had to say why they were all similar./

T. Which one? No. 14?

S. Yes, 14. /

S. There's only one that you can, you know, that you pick out why they're similar because / they're sort of, you know, other plants-you could get it mixed up with other plants otherwise .... /

S. Yes. They're all green.

T. Yes.

S. You know the cell...

T. There's something about those, / some structure

S. (Interjecting) Shiny.

T. ... in common that puts them into the one class. / Now we've been looking at structural features, haven't we?

S. Yes.

T. Come and have a look again, a minute. /

T. Now, they've been selected, / obviously, you see, because they've got the things on them, haven't they?

S. Yes. / They all produce a flower.

T. Yes, they all produce a flower. It's as simple as that. (40)

S. Yes, but they're not the only ones, they're just the common ones. /

T. Now, are the leaves the same?

S. Yes. They're all waxy. /

T. Are they all the same shape?

S. No ... prickles ....

T. They all have leaves for sure, / but you can get other plants such as your pine which has leaves as well, / but does it produce flowers?

S. No.

T. Has this thing produced a flower? / How do you know Fiona?

S. 'Cause that's the seed in the bottom there.

T. The fruit follows. / And what about these things, Mary?

S. The thick ones? /

T. We've got the thickest one, this one is opened, you can see where the seed was. So the seed pods must have come from? /

S. Flowers.

T. Flowers. So they all belong to a big group of plants known as the ...? /

T. Flowering Plants. Flowering Plants. Now do we find anything here that might / indicate that they belong to that group, the angiosperms. / They're all called angiosperms.

S. Yes. (41)

T. See, you haven't got the information about roots and whatnot, you know from past experience that they do have roots, but they track through the taxonomic key and they all belong / to the angiosperms because they have flowers. As distinct from the gymnosperms: / seeds between scales of a woody cone. Now what would be an example of a gymnosperm? / Just one along the bench there. What's that tree? /

S. (Inaudible answer)

T. (Inaudible) ... grow like that? /

S. What, that fern?

T. No, the next one down with the cones. /

S. Oh pine.

S. Pine.

T. Pine, yes pine. The pines are coniferous, mm! They all belong to the gymnosperms. / (Moves to another group of students)

T. Now, have you nearly finished, have you been right round?

S. Yes. /

T. Right, well get a sheet from the back, and you can use the books that are at the spots; / where the number is. // (Moves to another group)

T. Now, what have we got here Fleur?

S. Oh! Little green things. (42)

T. Mm! What shape would you say?

S. Circular

T. Mm mm! /

S. Little irregular brown ones.

T. Mm mm! What's the predominant organism there? /

S. Um!

T. What's present in the greatest quantity?

S. (Inaudible response) /

T. Ah ah! Now this is the little plant, where's the lemon branch? Pam, / could I see that twig of lemon? Now, what is under the microscope has come from this green stuff / here, this fine green stuff. It's been scraped off and a preparation made of it. / So how would you describe that plant. Any roots or limbs? / Each one of those makes a little single plant or alga. /

S. Single cell.

T. Good, Anything else? /

S. (Inaudible answer)

T. Mm! Green in colour, / what else? Does it produce flowers or roots or any other things?

S. No roots. /

T. No. So make a list of all the things it doesn't have and see if you can identify it. (43) (Teacher moves away)

S. I've only got sixteen: One, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fifteen, sixteen. /

S. That's because you've done some sort of ... (Inaudible).

S. Oh, that's for homework of course. /

S. Yes, That's so, don't worry.

T. Right, well if you've been the round, / get your sheet from the back of the room.

S. Where's that?

T. Typed sheet on the tray at the back of the room. / And answer the questions. You've discussed them but I want the answers written into the spaces on the sheet. /

S. Oh, it's different here, isn't it? It's funny. / (Teacher moves to assist student with microscope)

T. Now, we'll move this a bit because the one I want you to see is / really that one. Oh, I see you're just having another look around, O.K. / Sorry.

S. Oh, there's a batch there / and it's all green with fur and there's another batch that's red. /

T. Do you think all these things growing on this bed would have any similar features? /

S. Yea. (44) /

T. How would they get their food do you think Ramona?

S. From the bread.

T. In what way?

S. They eat it. /

T. Well, they don't have mouths or anything. So, how do they get it? /

S. They take the moisture out of it and they made a yeast ... / out of it.

S. They feed off it.

T. Right, / but in our digestive system, we have substances called enzymes which break down our meat and our carbohydrates / into things that the body can assimilate. / And the moulds, from those long thread-like structures, secrete enzymes into the bread, which digest the bread / into a liquidy sort of formula, and this diffuses back into the mould, and that uses it. / You notice a mouldy orange or apple; it's always slushy and mushy underneath. Well, this is the mould producing enzymes / that liquify that material of the fruit and this goes back into the mould / for its own food.
Now, do you think that green colouration is chlorophyll? (45)

S. The green one?

T. Mm! See, they're green ones there, aren't there? /

S. Yea.

T. Do you think that's chlorophyll? /

S. Oh, I don't know, I don't think so.

T. You don't think so? What's the function of chlorophyll? And then have a think. /

S. (Teachers' name), are they craters?

T. Well, they look like craters Jane don't they. They look like little mm! / saucer-shaped structures, or bowl-shaped structures. Do you think they might have any function? / Because there are quite a lot of them, aren't there?

S. They might be the little baby ones of it, could they? /

T. Could be concerned with reproduction, yes. Any roots, or leaves, or stems, / or what have you, in that one?

S. No.

T. No.

S. So what shall I call the, little craters? /

T. Yes. Call them anything you like so long as you know what you're talking about. And other people can understand you. // (Teacher moves on to another group).

T. Did you have a look at this one Helen? /

S. Oh, look (Inaudible) ... bottom. (46)

T. What do you think those are?

S. Seeds.

T. They're not seeds but they are / organs of reproduction. This thing doesn't produce any flowers, does it?

S. No. /

T. You know from your own experience; you know that ferns don't produce flowers. That got roots? /

S. No.

T. Leaves?

S. No.

T. Stems?

S. No.

T. Now, what about these? /

S. Supposed to be reconstructions for us.

T. Mm! They're little - spore cases they're called, / they're full of tiny spores. So they - You know, it's rather a complicated life-cycle, the fern; that's part of the life-cycle. / (Teacher moves to another student)

T. Now what about your conifer, do you notice these bits? /

S. No flowers.

S. Yes. Take the pollen out of that. /

T. So what part of the plant are they if they supply pollen? /

T. This is sexual reproduction isn't it. So what's that bit, is it the male or the female, / or is that the male-female? Both appearing on the same tree that you can see here. / Which part of the plant supplies the pollen; (47) the male or female? The seeds grow in that bit. /

S. So that must be the male.

T. That's the male, this is the female part of it. Well, there are two quite distinct structures on the one plant for reproduction. /

S. Would you need seeds for this?

T. Have you tried them? /

S. I have.

T. Are they a bit hard?

S. They're little black things like passion fruit seeds. /

T. Mm mm!

S. (Inaudible comment) Just look ....

T. Do they have any particular taste? /

S. Oh, I don't know really, I haven't had them for years. When we used to live up the bush we used to get lots of the stuff. /

T. Anyway that's part of the leaves of it? / Are these leaves?

S. Needle-like.

T. Mm mm! Now, have you nearly finished? /

T. Well, get a sheet from the back of the room and get on with the questions. // (Teacher moves on)

S. What does this (Inaudible) ... do?

S. From the bread. /

T. But how? Is it passed on?

S. Yes, through the stem.

T. How? (48)

S. I don't know (Laughter) /

S. Are these the ones that go with that? (Referring to question sheets)

T. You can look at this one. No, wait a minute. No, this one here / Those are extras.

S. Oh I see, that No. 55 is it? /

S. That's the bread one. /

S. The food just passing through the ... ?

T. Yes, it does, the food does, but how does the mould / prepare the food?

S. Sort of secrete it?

T. Mm mm! / Do you know what we call those substances that the mould secretes into the bread? /

S. Digested food.

T. Not digested food, but digestive ...?

S. Fluid. /

T. Yes, but what do we call these substances? We produce a lot of them in our own body? /

S. Bacteria?

S. They're like viruses or ... /

T. Yes, but what ....?

S. Amino acids.

T. No.

S. No. /

S. Seeds.

T. What do you call them?

S. What does it begin with?

T. En, EN (49)

S. Oh, I know it.

S. Enzymes.

T. Right, / that's right. They secrete those enzymes into the bread which digests the bread and makes it mushy; / or if it's a fruit makes it mushy and the liquid material goes back into the mould. /

S. (Inaudible comment)

S. Which sucks up the enzymes.

T. Well, it doesn't really suck, it's a matter of diffusion you know. / When the material's made mushy the liquid material diffuses back into the mould. /// (Teacher moves to another student)

T. What part of the plant do you think this is? /

S. (Inaudible answer)

T. Yes.

S. (Inaudible) ... mushroom ...? /

T. Well now, do mushrooms have flowers?

S. Well, the reproductory ...?

T. Well, it's not a flower, but it's a reproductive plant, all right. / What do you call these masses of long things?

S. Spores? /

T. No, they're not spores, but what do you call these? They look anything like a part of a fish? (50)

S. Ah! Gills.

T. You've got the gills, that's a descriptive word. /

S. Just here?

T. Mm! They're all the gills there. Have a look under the stereo, you can see what the gills look like under the stereo. In amongst those are masses of spores. / Any roots or leaves or ...?

S. These roots or not? /

T. Well this has just come out of the ground.

S. Not really.

T. No, not that you can see. / No. Right. No leaves.

S. No leaves.

T. Only a stem on this. /

S. ... mushroom would / hold all the spores, wouldn't it?

T. Mm mm! That's the reproductive bit. // (Moves to another student)

T. Now, what can you see under here? / Robyn.

S. They're green. Like one of those .... /

T. Is it moving? 'Cause if it were it would be moving / rapidly around wouldn't it? Just let me have a look to see if someone hasn't moved the slide. (51) / (Examines microscope) There's another bit there. Have a look there. / Focus with your fine adjustment. / What does it look like? /// (Student examines specimen under microscope)

S. Fine threads.

T. Mm mm! Now what do you think those sort of individual bits of a chain would be? // I'll put it on high power for you. /// (Teacher adjusts

microscope) I think you can see this bit right down the middle. (52) Just use your fine adjustment, and you'll see that the long thread is made up of - what do you think you'd call them? //
What do we call the tiny structures that make up bigger tissues / of living things and ...?

S. (Inaudible answer) /

T. Well, what's the nucleus in, what's the little structure that would contain a nucleus? /

S. Cell.

T. Good, yes. Lots of cells joined together, / so that single cells joined together and - can you see any relationship / between what you see under the microscope and what's in there?

S. No. /

T. Have a careful look. Now, you'll see that this is made up of what? / Thready material clumped together. Now when you spread that thready material / out, this is what you see. The long threads. They tend to get clumped together in the culture. /

S. What are (Inaudible) ...?

T. No, can't see one. (53) And I think this has come out of a pond actually down in the creek./ Some of them attach - and get stuck on to rocks, they don't have a holdfast like your seaweed. /

S. How does it stay up?

T. It gets tangled up and gets stuck on small prominences of rock, / so that it doesn't actually hook on to anything. / (Moves on to another specimen)
You had a look at this one Robyn? You see this? /

S. What is it?

T. Well, I'm not going to tell you because you're supposed to find out, aren't you? /

S. Are they little flowers or something? /

T. Well, they're something aren't they. What do you think they might grow into?

S. They might grow into leaves, / more leaves.

T. Yes, there was a structure here which I think has been lost - there's only one. / These little bits do grow up into little stalk-structures, there's one beginning to grow, it's beginning to grow a little stalk-structure. /

S. Yes.

T. They stay rather - (examining specimen) a green knob on top of the little stalk structure. /

S. That one? (54)

T. No. But besides those stalk structures can you see anything else on these flat things? Use your lens and see if you can see anything else.

S. Little spots. /

T. There's a good spot there, what's that thing? /

S. It's got little .... /

T. What's in there? With your lens Gary can you see what's in those little cup-shaped structures? /

S. Little green things.

T. Mm mm! What do you think they might be?

S. More leaves. /

T. Mm! Here's a good one, have a look at that one Jamie. / (Announcement over the school loudspeaker system) x/x/x/x/x/

T. What do you think these little bits might grow into eventually? (55)

S. New plants? /

T. More new plants, yes. So it's a reproductive structure, isn't it? These other little ones that are growing like this, / the knob-like ones are also reproductive structures. The important thing is, do you think this is a real leaf? /

S. No.

T. It's a leaf-like structure isn't it. /

S. No leaves.

T. Mm! Any roots?

S. Yea, there are little - little fine ones. Very fine ones. /

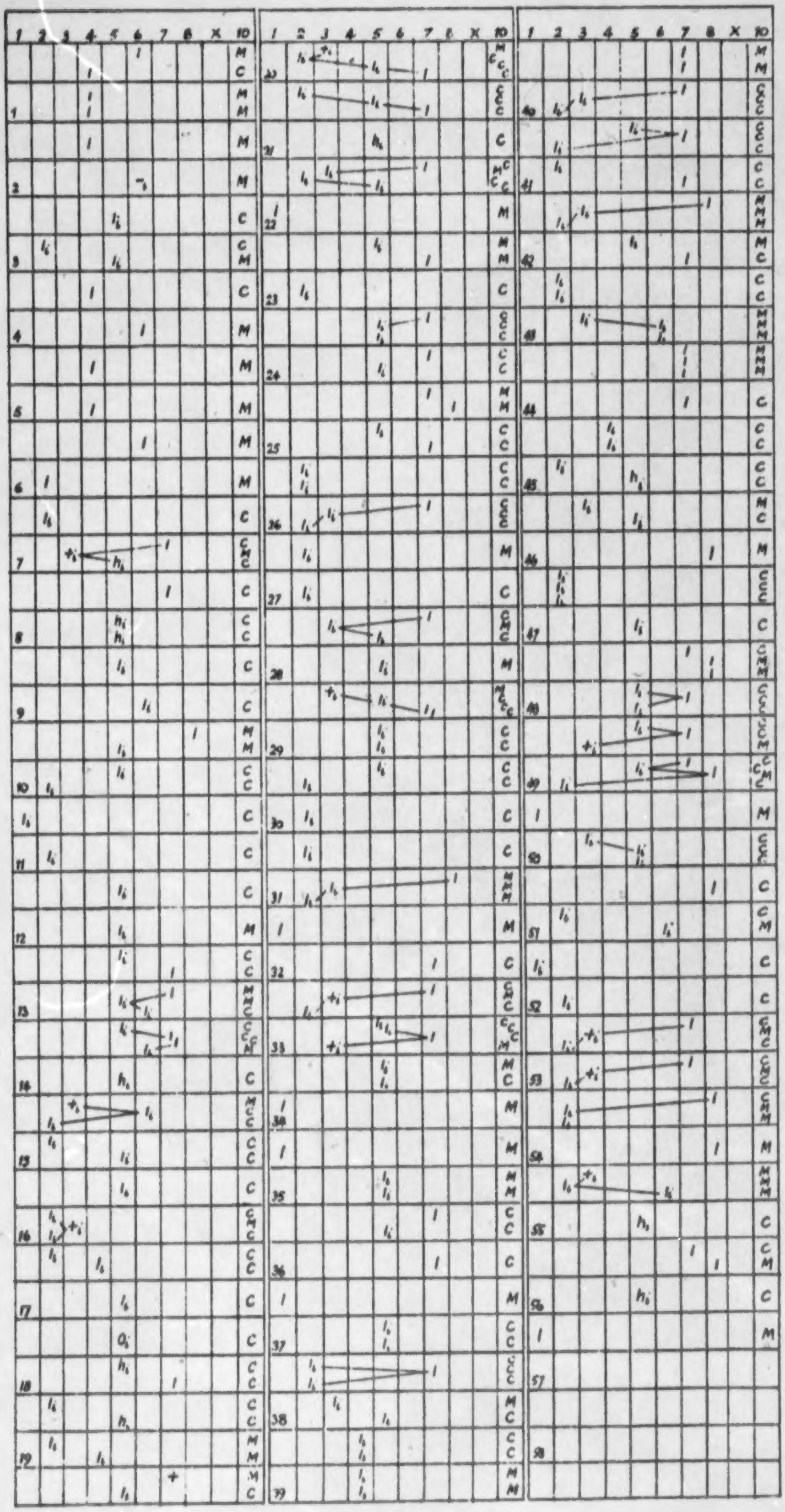
T. Mm mm! Very sort of fine roots, good. Anything else? Flowers, any flowers? /

S. Not unless they're flowers. So I'll just put it's got no flowers. /x/x/x/x/ (End of announcement. Siren signals end of lesson).

T. (Reading student notes) What makes you say it's a blue-green algae? (56)

T. Now, could I have your attention everybody please. // Just stand still for a moment. We'll finish this work off next time, / you should have finished most of your observations, and we'll answer our question sheet mainly from the literature / that we've got available. Leave the manuals in the centre of the benches would you please, and just move off. //

Appendix III. One-sixth sample coding of Lesson 21



REFERENCE NOTES

1. Evans, T.A. Flanders system of interaction analysis and science teacher effectiveness. Corvallis, Oregon : Oregon State University, 1970. (Mimeographed)
2. Flanders, N.A. Personal communication, December 2, 1970.
3. Hunt, D.E. Teachers' adaptation to students: from implicit to explicit matching. Toronto : Ontario Institute for Studies in Education, 1974. (Mimeographed)
4. Lees, R.J. Personal communication, approximately 1956.
5. Lucas, A.M. Personal communication, August 15, 1973.
6. Mackay, L.D. Test of cognitive preferences : Biology. Unpublished manuscript, Monash University, Australia, 1972.
7. Tamir, P. Personal communication, July 22, 1976.

REFERENCE LIST

Abramson, T., & Kagen, E. Familiarization of content and different response modes in programmed instruction. Journal of Educational Psychology, 1975, 67, 83-88.

Allon, N.R. Systems of classroom interaction analysis: A discussion of structural limitations. Journal of Experimental Education, 1969, 38(2), 1-4.

Allport, G.W. Pattern and growth in personality. London: Holt, Rinehart, & Winston, 1963.

Allport, G.W., & Odbert, H.S. Traits-names a psycholexical study. Psychological Monographs, 1936, 47, 1-171.

Amidon, E.J., & Flanders, N.A. The effects of direct and indirect teacher influence on dependent-prone students learning geometry. Journal of Educational Psychology, 1961, 52, 286-291.

Amidon, E.J. and Hough, J.B. (Eds.). Interaction analysis: theory, research and application. Reading, Mass.: Addison-Wesley, 1967.

Amidon, E., & Hunter, E. Improving teaching. The analysis of classroom verbal interaction. New York: Holt, Rinehart, & Winston, 1967.

Anderson, H.H., & Brewer, H.M. Studies of teachers' classroom personalities, I. Dominative and socially integrative behavior of kindergarten teachers. Applied Psychology Monograph No. 6 (APA), Stanford, Calif.: Stanford University Press, 1945.

Anderson, H.H., & Brewer, J.E. Studies of teachers' classroom personalities, II. Effects of dominative and integrative contacts on children's classroom behavior. Applied Psychology Monograph, No. 8 (APA), Stanford, Calif.: Stanford University Press, 1946. (a)

Anderson, H.H., & Brewer, J.E. Studies of teachers' classroom personalities, III. Follow-up studies of the effects of dominative and integrative contacts on children's behavior. Applied Psychology Monograph, No. 11 (APA), Stanford, Calif.: Stanford University Press, 1946. (b)

Anderson, R.C. Individual differences and problem solving. In R.M. Gagné (Ed.), Learning and individual differences, Columbus, Ohio: Merrill, 1967, pp.66-89.

Atkin, J.M., & Burnett, R.W. Science education. In R.L. Ebel (Ed.), Encyclopaedia of educational research (4th ed.), New York: The Macmillan Co., 1969, pp.1192-1206.

Atwood, R.K. A cognitive preference examination using chemistry content. Journal of Research in Science Teaching, 1967, 5, 34-35. (a)

Atwood, R.K. CHEM Study achievement among groups classified by cognitive preference scores. Journal of Research in Science Teaching, 1967, 5, 154-159. (b)

Atwood, R.K. Changes in cognitive preferences of high school physics students. School Science and Mathematics, 1969, 69, 697-699.

Australian Academy of Science. Biological Science: The Web of Life. Student's Manual. (2 pts.) Canberra, A.C.T.: Author, 1969.

Australian Academy of Science. Biological Science: The Web of Life. Teacher's Guide. (2 pts.) Canberra, A.C.T.: Author, 1969-1976.

Australian Council for Educational Research (ACER). ACER Advanced Test B40. Manual. Hawthorn, Victoria: Author, no date.

Australian Council for Educational Research (ACER). Biology Diagnostic and Achievement Tests. Teacher's Handbook. Hawthorn, Victoria: Author, 1972.

Ausubel, D.P. Some psychological aspects of the structure of knowledge. In S. Elam (Ed.), Education and the Structure of Knowledge. Chicago: Rand McNally, 1964, pp.221-248.

Ausubel, D.P. An evaluation of the 'Conceptual Schemes' approach to science curriculum development. Journal of Research in Science Teaching, 1965, 3, 255-264.

Ausubel, D.P. An evaluation of the BSCS approach to high school biology. The American Biology Teacher, 1966, 28, 176-186.

Ausubel, D.P. Educational psychology: A cognitive view. New York: Holt, Rinehart, & Winston, 1968.

Baird, L.L. Teaching styles: an exploratory study of dimensions and effects. Journal of Educational Psychology, 1973, 64, 15-21.

Bales, R.F. Interaction process analysis. Cambridge, Mass.: Addison-Wesley, 1950.

Balzer, A.L. Non-verbal and verbal behavior of biology teachers. The American Biology Teacher, 1969, 31, 226-229.

Barnett, H.C. An investigation of relationships among biology achievement, perception of teacher style, and cognitive preferences. Journal of Research in Science Teaching, 1974, 11, 141-147.

Belanger, M. Learning studies in science education. Review of Educational Research, 1969, 39, 377-395.

Bellack, A.A., Kliebard, H.M., Hyman, R.T., & Smith, F.L. Jr. The language of the classroom. New York: Teachers College Press, Columbia University, 1966.

Bennett, N. Teaching styles and pupil progress. London: Open Books, 1976.

Bennett, E.M., Alpert, R., & Goldstein, A.C. Communications through limited response questioning. *Public Opinion Quarterly*, 1954, *18*, 303-308.

Berliner, D.C., & Cahen, L.S. Trait-treatment interaction and learning. In F. Kerlinger (Ed.), *Review of research in education*, *1*. Itasca, Illinois: F.E. Peacock Publishers, Inc., 1973, pp.58-94.

Bernstein, B. Social structure, language, and learning. *Educational Research*, 1961, *3*, 163-176.

Bingman, R.M. (Ed.). *Inquiry objectives in the teaching of biology*. Kansas City, Mo: Mid-continent Regional Educational Laboratory, 1969.

Bingman, R.M. *The inquiry role approach as an instructional methodology* (A preliminary report). Kansas City, Mo.: Mid-continent Regional Educational Laboratory, June, 1970.

Block, J.H. (Ed.). *Mastery learning: Theory and practice*. New York: Holt, Rinehart, & Winston, 1971.

Bloom, B.S. Learning for mastery. *Education Comment* Vol. 1, No. 2, Center for the Study of Evaluation of Instructional Programs, UCLA, May, 1968.

Bloom, B.S., Englehart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. *Taxonomy of educational objectives. Handbook 1: Cognitive domain*. New York: Longmans Green & Co., 1956.

Boeck, C.H. The relative efficiency of reading and demonstration methods of instruction in developing scientific understandings. *Science Education*, 1956, *40*, 92-97.

Boleratz, J.M. Learning by discovery: An experimental study to measure its effectiveness for teaching value concepts. *Journal of Experimental Education*, 1967, *36*(2), 13-21.

Bonner, J.T. *The ideas of biology*. London: Eyre & Spottiswoode, 1963.

Bracht, G.H. Experimental factors related to aptitude-treatment interactions. *Review of Educational Research*, 1970, *40*, 627-645.

Bracht, G.H., & Glass, G.V. The external validity of experiments. *American Educational Research Journal*, 1968, *5*, 437-474.

Brandwein, R.F., Watson, F.G., & Blackwood, P.E. *Teaching high school science: A book of methods*. New York: Harcourt Brace, 1958.

Brazil, D. Kinds of English: Spoken, written, literary. *Educational Review*, 1969, *22*, 78-92.

Brewer, J.K. On the power of statistical tests in the 'American Educational Research Journal'. *American Educational Research Journal*, 1972, *9*, 391-401.

Bronfenbrenner, U. The experimental ecology of education. *Educational Researcher*, 1976, *5*(9), 5-15.

Bruner, J.S. *The process of education*. Cambridge, Mass.: Harvard U.P., 1960.

Bruner, J.S. Liberal education for all youth. *The Science Teacher*, 1965, *32*(1), 19-21.

Bruner, J.S. *Toward a theory of instruction*. Cambridge, Mass.: Harvard University Press, 1966.

Brunswik, E. *Perception and the representative design of psychological experiments*. Berkeley, Calif.: University of California Press, 1956.

Campbell, J.R. A longitudinal study in the stability of teachers' verbal behaviour. *Science Education*, 1972, *56*, 89-96.

Campbell, D.T., & Stanley, J.C. Experimental and quasi-experimental designs for research on teaching. In N.L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally, 1963, pp.171-246.

Carlson, J.S., & Ryan, F.L. Levels of cognitive functioning as related to anxiety. *Journal of Experimental Education*, 1969, *37*(4), 17-20.

Carroll, J.B. A model of school learning. *Teachers College Record*, 1963, *64*, 723-732.

Castaneda, A., McCandless, B.R., & Palermo, D.S. The children's form of the Manifest Anxiety Scale. *Child Development*, 1956, *27*, 317-326.

Cattell, R.B. *Personality and motivation. Structure and measurement*. New York: Harcourt, 1957.

Cattell, R.B. Guest editorial: Multivariate behavioral research and the integrative challenge. *Multivariate Behavioral Research*, 1966, *1*, 4-23.

Cattell, R.B., Beloff, H., & Coan, R.W. *Handbook for the IPAT High School Personality Questionnaire*. Champaign, Illinois: Institute for Personality and Ability Testing, 1958.

Cattell, R.B., & Coulter, M.A. Principles of behavioural taxonomy and the mathematical basis of the taxonome computer program. *The British Journal of Mathematical and Statistical Psychology*, 1966, *19*, 237-269.

Cohen, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 1960, *20*, 37-46.

Cohen, J. Multiple regression as a general data-analytic system. *Psychological Bulletin*, 1968, *70*, 426-443.

Cohen, J. Statistical power analysis for the behavioral sciences. New York: Academic Press, 1969.

Comrey, A.L. The minimum residual method of factor analysis. Psychological Reports, 1962, 11, 15-18.

Comrey, A.L., & Ahumada, A. An improved procedure and program for minimum residual factor analysis. Psychological Reports, 1964, 15, 91-96.

Coombs, C.H. On the use of inconsistency of preferences in psychological measurement. Journal of Experimental Psychology, 1958, 55, 1-7.

Coombs, C.H. A theory of data. New York: Wiley, 1964.

Coombs, J.R. The logic of teaching. In R.T. Hyman (Ed.), Teaching: Vantage points for study. Philadelphia: Lippincott, 1968, pp.420-427. (Reprinted from The High School Journal, 1966, 50, 22-29.)

Coop, R.H., & Brown, L.D. Effects of cognitive style and teaching method on categories of achievement. Journal of Educational Psychology, 1970, 61, 400-405.

Cornfield, J., & Tukey, J.W. Average values of mean squares in factorials. The Annals of Mathematical Statistics, 1956, 27, 907-949.

Coulter, J.C. The effectiveness of inductive laboratory, inductive demonstration and deductive laboratory in biology. Journal of Research in Science Teaching, 1966, 4, 185-186.

Cox, F.N. Correlates of general and test anxiety in children. Australian Journal of Psychology, 1960, 12, 169-177.

Crandall, V.J. Achievement. In Sixty-second Yearbook (Part 1) of the National Society for the Study of Education. Chicago: University of Chicago Press, 1963, pp.416-459.

Cronbach, L.J. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16, 297-334.

Cronbach, L.J. The two disciplines of scientific psychology. American Psychologist, 1957, 12, 671-684.

Cronbach, L.J. The logic of experiments on discovery. In L.S. Shulman & E.R. Keislar (Eds.), Learning by discovery: A critical appraisal. Chicago: Rand McNally, 1966, pp.77-92.

Cronbach, L.J. How can instruction be adapted to individual differences? In R.M. Gagné (Ed.), Learning and individual differences. Columbus, Ohio: Merrill, 1967, pp.23-39.

Cronbach, L.J. Essentials of psychological testing (3rd ed.). New York: Harper & Row, 1970.

Cronbach, L.J., & Gleser, G.C. Psychological tests and personnel decisions (2nd ed.). Urbana: Uni. Illinois Press, 1965.

Cronbach, L.J., Gleser, G., & Rajaratnam, N. Dependability of behavioral measurements. New York: Wiley, 1972.

Cronbach, L.J., & Snow, R.E. Final Report: Individual differences in learning ability as a function of instructional variables. Stanford, Calif.: School of Education, Stanford University, March, 1969. (ERIC Document Reproduction Service No. ED-029-001.)

Cunningham, W.G. The impact of student-teacher pairings on teacher effectiveness. American Educational Research Journal, 1975, 12, 169-189.

Daniels, R.L., & Stevens, J.P. The interaction between the internal-external locus of control and two methods of college instruction. American Educational Research Journal, 1976, 13, 103-113.

Darlington, R.B. Multiple regression in psychological research and practice. Psychological Bulletin, 1968, 69, 161-182.

Davids, A. Cognitive styles in potential scientists and underachieving high school students. Journal of Special Education, 1968, 2, 197.

Davis, O.L., & Tinsley, D.C. Cognitive objectives revealed by classroom questions asked by social studies student teachers. In R.T. Hyman (Ed.), Teaching: Vantage points for study. Philadelphia: Lippincott, 1968, pp.140-146. (Reprinted from Peabody Journal of Education, 1967, 45, 21-26.)

Dershimer, R.A., & Iannaccone, L. Social and political influences on educational research. In R.M.W. Travers (Ed.), Second handbook of research on teaching. Chicago: Rand McNally, 1973.

Di Vesta, F.J. Theory and measurement of individual differences in studies of trait by treatment interaction. Educational Psychologist, 1973, 10, 67-75.

Domino, G. Differential predictions of academic achievement in conforming and independent settings. Journal of Educational Psychology, 1968, 59, 256-260.

Doty, B.A. Teaching method effectiveness in relation to certain student characteristics. Journal of Educational Research, 1967, 60, 363-365.

Doty, B.A., & Doty, L.A. Programed instructional effectiveness in relation to certain student characteristics. Journal of Educational Psychology, 1964, 55, 334-338.

Dubois, E.A.C. Induction and deduction (Doctoral dissertation, Harvard University, 1966). Dissertation Abstracts, 1966, 28, 492A (University Microfilms No. 67-5563).

Dunkin, M.J., & Biddle, B.J. The study of teaching. New York: Holt, Rinehart, & Winston, 1974.

Ebel, R.L. The relation of item discrimination to test reliability. *Journal of Educational Measurement*, 1967, *4*, 125-128.

Ebel, R.L. The value of internal consistency in classroom examinations. *Journal of Educational Measurement*, 1968, *5*, 71-73.

Egelston, J. Inductive vs. traditional methods of teaching high school biology laboratory experiments. *Science Education*, 1973, *57*, 467-477.

Ehrlich, H.J. Dogmatism and learning. *Journal of Abnormal and Social Psychology*, 1961, *62*, 148-149.

Elashoff, J.D. Analysis of covariance: A delicate instrument. *American Educational Research Journal*, 1969, *6*, 383-401.

Emmer, E.T. Direct observation of classroom behavior. *International Review of Education*, 1972, *18*, 473-489.

Emmer, E.T., & Peck, R.F. Dimensions of classroom behavior. *Journal of Educational Psychology*, 1973, *64*, 223-240.

Endler, N.S., Hunt, J.McV., & Rosenstein, A.J. An S-R inventory of anxiousness. *Psychological Monographs*, 1962, *76*, (17, Whole No. 536)

Entwistle, N.J., & Cunningham, S. Neuroticism and school attainment - A linear relationship? *British Journal of Educational Psychology*, 1968, *38*, 123-132.

Evans, T.P. A category system for teacher behaviors. *The American Biology Teacher*, 1969, *31*, 221-225.

Eysenck, H.J. *The structure of human personality*. London: Methuen, 1953.

Eysenck, H.J. *Fact and fiction in psychology*. London: Penguin, 1965.

Eysenck, H.J. *The biological basis of personality*. Springfield: Charles C. Thomas, 1967.

Eysenck, H.J., & Cookson, D. Personality in primary school children: 1 - Ability and achievement. *British Journal of Educational Psychology*, 1969, *39*, 109-122.

Eysenck, H.J., & Eysenck, S.B.G. 'Psychoticism' in children: a new personality variable. *Research in Education*, 1969, *1*, 29-37.

Eysenck, S.B.G., & Eysenck, H.J. The validity of questionnaires and rating assessments of extraversion and neuroticism and their factorial validity. *British Journal of Psychology*, 1963, *54*, 51-62.

Fillenbaum, S., & Jackson, A. Dogmatism and anxiety in relation to problem solving: An extension of Rokeach's results. *Journal of Abnormal and Social Psychology*, 1961, *63*, 212-214.

Fischler, A.S., & Anastasiow, N.J. In-service education in science (A pilot). Journal of Research in Science Teaching, 1965, 3, 280-285.

Fisher, R.A. Statistical methods for research workers. Edinburgh: Oliver & Boyd, 1925.

Flanders, N.A. Teacher influence, pupil attitudes, and achievement. In R.T. Hyman (Ed.), Teaching: Vantage points for study. Philadelphia: Lippincott, 1968, pp.251-265. (Reprinted from USOE Cooperative Research Monograph No. 12. Washington, D.C.: USOE, 1965.)

Flanders, N.A. The problems of observer training and reliability. In E.J. Amidon & J.B. Hough (Eds.), Interaction analysis: Theory, research and application. Reading, Mass.: Addison-Wesley, 1967, pp.158-166.

Flanders, N.A. Analyzing teaching behavior. Reading, Mass.: Addison-Wesley, 1970.

Flanders, N.A., & Simon, A. Teacher effectiveness. In R.L. Ebel (Ed.), Encyclopaedia of educational research. Toronto: Macmillan, 1969, pp.1423-1437.

Franks, C.M. Personality factors and the rate of conditioning. British Journal of Psychology, 1957, 48, 119-126.

Furneaux, W.D. Some psychometric characteristics of university students seeking psychiatric help. Bulletin of the British Psychological Society, 1956, 29, 7.

Gagné, R.M. The learning requirements for enquiry. Journal of Research in Science Teaching, 1963, 1, 144-153.

Gagné, R.M. (Ed.) Learning and individual differences. Columbus: Charles E. Merrill, 1967.

Gagné, R.M. Learning hierarchies. Educational Psychologist, 1968, 6, 1-9.

Gagné, R.M. The conditions of learning (2nd ed.). New York: Holt, Rinehart, & Winston, 1970.

Gallagher, J.J. Teacher variation in concept presentation in BSCS curriculum program. BSCS Newsletter, No. 30, Jan. 1967.

Gallagher, J.J., & Aschner, M.J. A preliminary report on analyses of classroom interaction. Merrill-Palmer Quarterly, 1963, 9, 183-194.

Galloway, C.M. Nonverbal communication in teaching. Educational Leadership, 1966, 24, 55-63.

Gardner, P.L. Attitudes to physics. Unpublished Ph.D. Thesis, Monash University, 1972.

Gardner, P.L. Attitude measurement: a critique of some recent research. Educational Research, 1975, 17(2), 101-109. (a)

Gardner, P.L. Attitudes to science: A review. Studies in Science Education, 1975, 2, 1-41. (b)

Gardner, P.L. Science and the structure of knowledge. In P.L. Gardner (Ed.), The structure of science education. Hawthorn, Victoria: Longman Australia, 1975, pp.1-40. (c)

Gardner, P.L. Scales and statistics. Review of Educational Research, 1975, 45, 43-47. (d)

Gaudry, E., & Spielberger, C.D. Anxiety and educational achievement. Sydney: John Wiley & Sons, 1971.

Getzels, J., & Madaus, G.F. Creativity. In R.L. Ebel (Ed.), Encyclopaedia of educational research. Toronto: Macmillan, 1969, pp.267-274.

Glaser, R. Variables in discovery learning. In L.S. Shulman & E.R. Keislar (Eds.), Learning by discovery: A critical appraisal. Chicago: Rand McNally, 1966, pp.13-26.

Glaser, R. Learning. In R.L. Ebel (Ed.), Encyclopaedia of educational research. Toronto: Macmillan, 1969, pp.706-733.

Glaser, R. Individuals and learning: The new aptitudes. Educational Researcher, 1972, 1(6), 5-13.

Glaser, R., & Nitko, A.J. Measurement in learning and instruction. In R.L. Thorndike (Ed.), Educational Measurement. Washington: American Council on Education, 1971, pp.625-670.

Glass, B. Renascent biology: a report on the AIBS Biological Sciences Curriculum Study. In R.W. Heath (Ed.), New curricula. New York: Harper & Row, 1964, pp.94-119. (Reprinted from The School Review, 1962.)

Gocka, E.F. Regression analysis of proportional cell data. Psychological Bulletin, 1973, 80, 25-27.

Gough, H.G. California Psychological Inventory. Palo Alto, Calif.: Consulting Psychologists Press, 1956.

Gray, J.L., & Kneif, L.M. Relationship between cognitive style and school achievement. Journal of Experimental Education, 1975, 43(4), 67-71.

Grimes, J.W., & Allinsmith, W. Compulsivity, anxiety and school achievement. Merrill-Palmer Quarterly, 1961, 7, 247-271.

Grobman, H. Development curriculum projects: Decision points and processes. Itasca, Illinois: Peacock Publishers, Inc., 1970.

Guilford, J.P. When not to factor analyze. Psychological Bulletin, 1952, 49, 26-37.

Guilford, J.P. Structure of intellect. Psychological Bulletin, 1956, 53, 267-293.

Hambleton, R.K., & Novick, M.R. Towards an integration of theory and method for criterion-referenced tests. Journal of Educational Measurement, 1973, 10, 159-170.

Hanson, N.R. Patterns of discovery: An inquiry into the conceptual foundations of science. London: Cambridge University Press, 1961.

Harvey, O.J., Hunt, D.E., & Schroder, H.M. Conceptual systems and personality organization. New York: Wiley, 1961.

Heath, R.W. Curriculum, cognition and educational measurement. Educational and Psychological Measurement, 1964, 24, 239-253.

Heist, P., & Yonge, G. Omnibus Personality Inventory, Form F, Manual. New York: The Psychological Corporation, 1962, 1968.

Herbert, J.D. A system for analyzing lessons. New York: Teachers College Press, 1967.

Hermann, G. Learning by discovery: A critical review of studies. Journal of Experimental Education, 1969, 38(1), 58-72.

Herron, J.D., Luce, T.G., & Neie, V.E. The proper experimental unit: Comparative analyses of empirical data. Journal of Research in Science Teaching, 1976, 13, 19-27.

Herron, M.D. The nature of scientific enquiry. School Review, 1971, 79, 171-212.

Holzman, P.S., & Gardner, R.W. Leveling-sharpening and memory organization. Journal of Abnormal and Social Psychology, 1960, 61, 176-180.

Horn, J.L. Some characteristics of classroom examinations. Journal of Educational Measurement, 1966, 3, 293-295.

Horton, R.E. Measurable outcomes of laboratory instruction in elementary college physics. American Journal of Physics, 1952, 20, 136-141.

Hough, J.B., & Duncan, J.K. Teaching: description and analysis. Reading, Mass.: Addison-Wesley, 1970.

Hovland, C.I. The generalization of conditioned responses I. The sensory generalization of conditioned responses with varying frequencies of tone. Journal of General Psychology, 1937, 17, 125-148.

Hughes, M.M. What is teaching? One viewpoint. Educational Leadership, 1962, 19, 251-259.

Humphreys, L.C., & Dachler, H.P. Jensen's theory of intelligence. Journal of Educational Psychology, 1969, 60, 419-426.

Hunt, D.E. Person-environment interaction: A challenge found wanting before it was tried. *Review of Educational Research*, 1975, *45*, 209-230.

Hurd, P.DeH. *Biological education in American secondary schools, 1890-1960*. BSCS Bulletin No. 1, Washington, D.C.: American Institute of Biological Science (AIBS), 1961.

Huynh, H. On consistency of decisions in criterion-referenced testing. *Journal of Educational Measurement*, 1976, *13*, 253-264.

Hyman, R.T. (Ed.). *Teaching: Vantage points for study*. Philadelphia: Lippincott, 1968.

Jensen, A.R. Patterns of mental ability and socio-economic status. *Proceedings of the National Academy of Sciences*, 1968, *60*, 1330-1337.

Joyce, B.R., & Harootunian, B. *The structure of teaching*. Chicago: Science Research Associates, 1967.

Kagan, J. Impulsive and reflective children. In J. Krumboltz (Ed.), *Learning and the educational process*. Chicago: Rand McNally, 1965, pp.131-161.

Kagan, J., & Kogan, N. Individual variation in cognitive processes. In P.H. Mussen (Ed.), *Carmichael's manual of child psychology*, Volume 1 (3rd ed.). New York: Wiley, 1970, pp.1273-1365.

Kagan, J., Moss, H.A., & Sigel, I.E. The psychological significance of styles of Conceptualization. *Basic cognitive processes in children. Monograph of the Society for Research in Child Development*, 1963, *28*(2), 73-112.

Kaiser, H.F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, *23*, 187-200.

Kaiser, H.F. An index of factorial simplicity. *Psychometrika*, 1974, *39*, 31-36.

Katahn, M. Interaction of anxiety and ability in complex learning situations. *Journal of Personal and Social Psychology*, 1966, *3*, 475-479.

Kemp, C.G. Effects of dogmatism on critical thinking. *School Science and Mathematics*, 1960, *60*, 314-319.

Kempa, R.F., & Dubé, G.E. Cognitive preference orientations in students of chemistry. *British Journal of Educational Psychology*, 1973, *43*, 279-288.

Kerlinger, F.N., & Pedhazur, E.J. *Multiple regression in behavioral research*. New York: Holt, Rinehart, & Winston, 1973.

Kim, J-O., & Kohout, F.J. Multiple regression analysis: subprogram regression. In N.H. Nie, C.H. Hull, J.G. Jenkins, K. Steinbrenner, & D.H. Brent, *Statistical package for the social sciences (SPSS)* (2nd ed.). New York: McGraw-Hill, 1975.

King, A.R., & Brownell, J.A. *The curriculum and the disciplines of knowledge*. New York: John Wiley & Sons, 1966.

Klein, S.P., Fredericksen, N., & Evans, F.R. Anxiety and learning to formulate hypotheses. *Journal of Educational Psychology*, 1969, *60*, 465-475.

Klinckmann, E. (Supervisor) *Biology teachers' handbook* (2nd ed.). New York: John Wiley & Sons, 1970.

Koay, P.L. *Cognitive preferences and achievement in mathematics*. Unpublished M.Ed. thesis, Monash University, 1975.

Krathwohl, D.R., Bloom, B.S., & Masia, B.B. *Taxonomy of educational objectives, Handbook II: Affective domain*. London: Longmans Green, 1964.

Kruglak, H. Achievement of physics students with and without laboratory work. *American Journal of Physics*, 1953, *21*, 14-16.

Kruglak, H. The measurement of laboratory achievement. *American Journal of Physics*, 1954, *22*, 442-462.

Kruglak, H. The measurement of laboratory achievement. *American Journal of Physics*, 1955, *23*, 82-87.

Krumboltz, J.D., & Yabroff, W.W. The comparative effects of inductive and deductive sequences in programed instruction. *American Educational Research Journal*, 1965, *2*, 223-235.

Kuder, G.F., & Richardson, M.W. The theory of the estimation of test reliability. *Psychometrika*, 1937, *2*, 151-160.

Levy, L.H. Anxiety and scientists' behavior. *American Psychologist*, 1961, *2*, 66-68.

Lewin, K. *Dynamic theory of personality*. New York: McGraw-Hill, 1935.

Lewis, W.W., Newell, J.M., & Withall, J. An analysis of classroom patterns of communication. *Psychological Report*, 1961, *9*, 211-219.

Likert, R. A technique for the measurement of attitudes. In M. Fishbein (Ed.), *Readings in attitude theory and measurement*. New York: Wiley, 1967. (Reprinted from *Archives of Psychology*, 1932, *140*, 44-53.

Lindquist, E.F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin Co., 1953.

Livingston, S.A. Criterion-referenced applications of classical test theory. *Journal of Educational Measurement*, 1972, *9*, 13-26.

Lord, F.M. Sampling error due to choice of split in split-half reliability coefficients. *Journal of Experimental Education*, 1956, *24*, 245-249.

Lord, F.M. Elementary models for measuring change. In C.W. Harris (Ed.), Problems in measuring change. Madison, Wisc.: University of Wisconsin Press, 1963.

Lubin, A. The interpretation of significant interaction. Educational and Psychological Measurement, 1961, 21, 809-818.

Lucas, A.M., & Broadhurst, N.A. Changes in some content-free skills, knowledge, and attitudes during two terms of Grade 12 biology instruction in ten South Australian schools. The Australian Science Teachers Journal, 1972, 18(1), 66-74.

Lumsdaine, A.A. Instruments and media of instruction. In N.L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963.

Lynn, R., & Gordon, I.E. The relation of neuroticism and extraversion to intelligence and educational attainment. British Journal of Educational Psychology, 1961, 31, 194-203.

Mackay, L.D. Changes in Victorian physics students during two years of physics study. Australian Physicist, July, 1970, 103-109.

Mackay, L.D. Changes in affective domain objectives during two years of physics study in Victorian schools. The Australian Science Teachers Journal, 1971, 17(1), 67-71. (a)

Mackay, L.D. Cognitive preferences and achievement in physics. Unpublished Ph.D. thesis, Monash University, 1971. (b)

Mackay, L.D. Changes in cognitive preferences during two years of physics study in Victorian schools. The Australian Science Teachers Journal, 1972, 18(1), 63-66.

Mackay, L.D. Cognitive preferences and achievement in physics, chemistry, science and mathematics. Research in Science Education (Australian Science Education Research Association), 1975, 5, 49-58.

Mager, R.F. Preparing instructional objectives. Palo Alto, Calif.: Fearon Publishers, 1962.

Maier, M.H. & Jacobs, P.I. The effects of variations in a self-instructional program on instructional outcomes. Psychological Reports, 1966, 18, 539-546.

Marks, R.L. CBA high school chemistry and concept formation. Journal of Chemical Education, 1967, 44, 471-474.

May, F.B., & De Vault, M.V. Hypothetical dimensions of teachers' communication. American Educational Research Journal, 1967, 4, 271-278.

Mayer, D.E. Information processing variables in learning to solve problems. Review of Educational Research, 1975, 45, 525-541.

McArthur, J.T., Fraser, B.J., & West, L.H.T. Some issues in the simulation of ANOVA by multiple-regression techniques. Research in Science Education, 1975, 5, 199-207.

McGaw, B., Wardrop, J.L., & Bunda, M.A. Classroom observation schemes: Where are the errors? American Educational Research Journal, 1972, 9, 13-27.

McKeachie, W.J. The decline and fall of the laws of learning. Educational Researcher, 1974, 3(3), 7-11.

McKeachie, W.J., Yi-Guang Lin, Milholland, J., & Isaacson, R. Student affiliation motives, teacher warmth, and academic achievement. Journal of Personality and Social Psychology, 1966, 4, 457-461.

McNaught, C.M. Cognitive preferences and achievement in chemistry. Unpublished M.Ed. thesis, Monash University, 1975.

McNemar, Q. Lost: Our intelligence? Why? American Psychologist, 1964, 19, 871-882.

Medley, D.M., & Mitzel, H.E. Application of analysis of variance to the estimation of the reliability of observations of teachers' classroom behaviour. Journal of Experimental Education, 1958, 27(1), 26-36.

Medley, D.M., & Mitzel, H.E. Measuring classroom behavior by systematic observation. In N.L. Gage (Ed.), Handbook of Research on Teaching. Chicago: Rand McNally, 1963, pp.247-328.

Meichenbaum, D.H., & Goodman, J. Reflection-impulsivity and verbal control of motor behavior. Child Development, 1969, 40, 785-797.

Messick, S. The criterion problem in the evaluation of instruction: Assessing possible not just intended outcomes. In M.C. Wittrock & D.E. Wiley (Eds.), The Evaluation of Instruction: Issues and Problems. Chicago: Holt, Rinehart, & Winston, 1970, pp.185-202.

Millman, J. In service of generalization. Psychology in the Schools, 1966, 3, 333-339.

Mitchell, J.V., Jnr. Education's challenge to psychology: The prediction of behavior from person-environment interactions. Review of Educational Research, 1969, 39, 695-721.

Mitzel, H.E. Teacher effectiveness. In C.W. Harris (Ed.), Encyclopaedia of educational research (3rd ed.). Toronto: Macmillan, 1960, pp.1481-1486.

Moon, I.C. A study of verbal behavior patterns in primary grade classrooms during science activities. Journal of Research in Science Teaching, 1971, 8, 171-177.

Morgan, D.G. (Ed.) Biological science: The web of life (2nd ed.). Canberra, A.C.T.: Australian Academy of Science, 1973.

Morrison, A., & McIntyre, D. Teachers and teaching. Harmondsworth: Penguin Books Ltd., 1969.

Moskowitz, G. The attitudes and teaching patterns of cooperating teachers and student teachers trained in interaction analysis. In E.J. Amidon & J.B. Hough (Eds.), Interaction analysis: Theory, research, and application. Reading, Mass.: Addison-Wesley, 1967, pp.271-282.

Mouw, J.T. The effect of dogmatism on levels of cognitive processes. Journal of Educational Psychology, 1969, 60, 365-369.

Murphy, G.W. Content versus process centered biology laboratories, part II: The development of knowledge, scientific attitudes, problem solving ability, and interest in biology. Science Education, 1968, 52, 148-162.

Neidt, C.O., & Hedlund, D.E. Longitudinal relationships between cognitive learning and affective learning outcomes. Journal of Experimental Education, 1969, 37(3), 56-60.

Nelson, L.N. (Ed.) The nature of teaching. Waltham, Mass.: Blaisdell Publishing Company, 1969.

Neville, D., Pfost, P., & Dobbs, V. The relationship between test anxiety and silent reading gain. American Educational Research Journal, 1967, 4, 45-50.

Newsom, R.S., Eischens, R., & Looft, W.R. Intrinsic individual differences: A basis for enhancing instructional programs. Journal of Educational Research, 1972, 65, 387-392.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, J., & Bent, D.H. Statistical package for the social sciences SPSS (2nd ed.). New York: McGraw-Hill, 1975.

Novak, J.D. What should we teach in biology? Journal of Research in Science Teaching, 1963, 3, 241-243.

Novak, J.D., Ring, D.G., & Tamir, P. Interpretation of research findings in terms of Ausubel's theory and implications for science education. Science Education, 1971, 55, 483-526.

Nuthall, G. An experimental comparison of alternative strategies for teaching concepts. American Educational Research Journal, 1968, 5, 561-584.

Nuthall, G. The search for ideal teaching methods. London Educational Review, 1973, 2(3), 29-37.

Nuthall, G., & Church, J. Observation systems used with recording media. International Review of Education, 1972, 18, 491-507.

Ogunyemi, E.L. Cognitive styles and student science achievement in Nigeria. Journal of Experimental Education, 1973, 42(1), 59-63.

Oliver, D.W., & Shaver, J.P. Teacher style and the analysis of the student-teacher dialogue. In R.T. Hyman (Ed.), Teaching: Vantage points for study. Philadelphia: Lippincott, 1968, pp.404-420. (Reprinted from Chapter 12 of The analysis of public controversy: A study in citizenship education. USOE, Cooperative Research Project No. 8145. Cambridge: Harvard Graduate School of Education, 1962.)

Ott, M.D., & Macklin, D.B. A trait-treatment interaction in a college physics course. Journal of Research in Science Teaching, 1975, 12, 111-119.

Overall, J.E., & Spiegel, D.K. Concerning least squares analysis of experimental data. Psychological Bulletin, 1969, 72, 311-322.

Overall, J.E., & Spiegel, D.K. Comment on "Regression analysis of proportional cell data". Psychological Bulletin, 1973, 80, 28-30.

Overall, J.E., Spiegel, D.K., & Cohen, J. Equivalence of orthogonal and nonorthogonal analysis of variance. Psychological Bulletin, 1975, 82, 182-186.

Parakh, J.S. A study of teacher-pupil interaction in high school biology classes. Part I. Journal of Research in Science Teaching, 1969, 6, 284-292.

Parakh, J.S. A study of teacher-pupil interaction in high school biology classes. Part II. Journal of Research in Science Teaching, 1968, 5, 183-192.

Peckham, P.D., Glass, G.V., & Hopkins, K.D. The experimental unit in statistical analysis: Comparative experiments with intact groups. Journal of Special Education, 1969, 3, 337-349.

Pella, M.O. The laboratory and science teaching. The Science Teacher, 1961, 28(5), 29-31.

Pella, M.O., & Sherman, J. A comparison of two methods of utilizing laboratory activities in teaching the course IPS. School Science and Mathematics, 1969, 69, 303-314.

Penick, J.E., & Brewer, J.K. The power of statistical tests in science teaching research. Journal of Research in Science Teaching, 1972, 9, 377-381.

Pereboom, A.C. Some fundamental problems in experimental psychology: An overview. Psychological Reports, 1971, 28, 439-455.

Perkins, H.V. A procedure for assessing the classroom behavior of students and teachers. American Educational Research Journal, 1964, 1, 249-260.

Piaget, J. The psychology of intelligence. London: Routledge, Kegan Paul, 1947.

Popham, W.J., & Husek, T.R. Implications of Criterion-referenced measurement. *Journal of Educational Measurement*, 1969, *6*, 1-9.

Posner, G.J., & Strike, K.A. A categorization scheme for principles of sequencing content. *Review of Educational Research*, 1976, *46*, 665-690.

Ramsey, G.A. Science teaching as an instructional system. In P.L. Gardner (Ed.), *The structure of science education*. Hawthorn, Vic.: Longmans Australia, 1975, pp.94-114.

Ramsey, G.A., & Howe, R.W. An analysis of research on instructional procedures in secondary school science, Part I: Outcomes of instruction. *The Science Teacher*, 1969, *36*(3), 62-70. (a)

Ramsey, G.A., & Howe, R.W. An analysis of research on instructional procedures in secondary school science, Part II: Instructional procedures. *The Science Teacher*, 1969, *36*(4), 72-81. (b)

Raths, J. The appropriate experimental unit. *Educational Leadership*, 1967, *25*, 263-266.

Raths, J.D., Pancella, J.R., & van Ness, J.S. (Eds.) *Studying teaching* (2nd ed.). Englewood Cliffs, N.J.: Prentice-Hall, 1971.

Rawlings, R.R., Jnr. Comments on the Overall and Spiegel paper. *Psychological Bulletin*, 1973, *79*, 168-169.

Rhetts, J.E. Attribute-treatment interactions and individualized instruction: A conceptual framework and an example from Project PLAN. In L. Sperry (Ed.), *Learning performance and individual differences*. Glenview, Ill.: Scott Foresman & Co., 1972, pp.269-285.

Ripple, R.E., Millman, J., & Glock, M.D. Learner characteristics and instructional mode: A search for disordinal interactions. *Journal of Educational Psychology*, 1969, *60*, 113-120.

Robinson, J.T. Science teaching and the nature of science. *Journal of Research in Science Teaching*, 1965, *3*, 37-50.

Robinson, J.T. Evaluating laboratory work in high school biology. *The American Biology Teacher*, 1969, *31*, 236-239.

Rokeach, M. The nature and meaning of dogmatism. *Psychological Review*, 1954, *61*, 194-204.

Rokeach, M. *The open and closed mind*. New York: Basic Books, 1960.

Rosenshine, B. The stability of teacher effects upon student achievement. *Review of Educational Research*, 1970, *40*, 647-622.

Rosenshine, B. *Teaching behaviours and student achievement*. IEA Studies No. 1. Windsor, Berks.: National Foundation for Educational Research in England and Wales, 1971.

Rosenshine, B. The smallest meaningful sample of classroom transactions. *Journal of Research in Science Teaching*, 1973, *10*, 221-226.

Rosenshine, B., & Furst, N. The use of direct observation to study teaching. In R.M.W. Travers (Ed.), *Second handbook of research on teaching*. Chicago: Rand McNally, 1973, pp.122-183.

Rowley, G.L. The reliability of observational measures. *American Educational Research Journal*, 1976, *13*, 51-59.

Ruebush, B.K. Anxiety. In *Sixty-second Yearbook (Part I) of the National Society for the Study of Education*. Chicago: University of Chicago Press, 1963, pp.460-516.

Rushton, J. The relationship between personality characteristics and scholastic success in eleven-year-old children. *British Journal of Educational Psychology*, 1966, *36*, 178-184.

Rutherford, F.J. The role of inquiry in science teaching. *Journal of Research in Science Teaching*, 1964, *2*, 80-84.

Sakmyser, D.D. Comparison of inductive and deductive programmed instruction on chemical equilibrium for high school chemistry students. *Journal of Research in Science Teaching*, 1974, *11*, 67-77.

Salomon, G. Heuristic models for the generation of aptitude-treatment interaction hypotheses. *Review of Educational Research*, 1972, *42*, 327-343.

Salomon, G. Internalization of filmic schematic operators in interaction with learner's aptitude. *Journal of Educational Psychology*, 1974, *66*, 499-511.

Sanders, N. *Classroom questions. What kinds?* New York: Harper & Row, 1966.

Sanford, N., Webster, H., & Freedman, M. Impulse expression as a variable of personality. *Psychological Monographs*, 1957, *71* (11, Whole No. 440).

Sarason, I.G. Effects on verbal learning of anxiety, reassurance, and meaningfulness of material. *Journal of Experimental Psychology*, 1958, *56*, 472-477.

Sarason, S.B., Davidson, K.S., Lighthall, F.F., Waite, R.R., & Ruebush, B.K. *Anxiety in elementary school children*. New York: Wiley, 1960.

Savage, R.D. Personality factors and academic performance. *British Journal of Educational Psychology*, 1962, *32*, 251-253.

Savage, R.D. Personality factors and academic achievement in junior school children. *British Journal of Educational Psychology*, 1966, *36*, 91-92.

Schefler, W.C. A comparison between inductive and illustrative laboratories in college biology. Journal of Research in Science Teaching, 1965, 3, 218-222.

Schmedemann, G., & La Shier, N.S. Cognitive preferences of students and selected characteristics of their PSSC teachers. Journal of Research in Science Teaching, 1967, 5, 40-42.

Schock, N.H. An analysis of the relationship which exists between cognitive and affective educational objectives. Journal of Research in Science Teaching, 1973, 10, 299-315.

Schroder, H.M. Conceptual complexity and personality organization. In H.M. Schroder & P. Suedfeld (Eds.), Personality theory and information processing. New York: Ronald, 1971, pp.240-274.

Schroder, H.M., Driver, M.J., & Streufert, S. Human information processing. New York: Holt, Rinehart, & Winston, 1967.

Schwab, J.J. The teaching of science as enquiry. In J.J. Schwab & P.F. Brandwein (Eds.), The teaching of science. Cambridge: Harvard U.P., 1962, pp.1-103.

Schwab, J.J. The structure of the natural sciences. In G.W. Ford & L. Pugno (Eds.), The structure of knowledge and the curriculum. Chicago: Rand McNally & Co., 1964, pp.31-49.

Scott, N.C., Jnr. Science concept achievement and cognitive functions. Journal of Research in Science Teaching, 1964, 2, 7-16.

Scott, N.C., Jnr. Cognitive style and inquiry strategy: A five-year study. Journal of Research in Science Teaching, 1973, 10, 323-330.

Scott, W.A. Reliability of content analysis: The case of nominal scale coding. Public Opinion Quarterly, 1955, 19, 321-325.

Sells, S.B. Personality. In R.L. Ebel (Ed.), Encyclopaedia of educational research. Toronto: Macmillan, 1969, pp.935-946.

Shavelson, R., & Dempsey-Atwood, N. Generalizability of measures of teaching behavior. Review of Educational Research, 1976, 46, 553-611.

Shulman, L.S., & Keislar, E.R. (Eds.) Learning by discovery: A critical appraisal. Chicago: Rand McNally, 1966.

Shulman, L.S., & Tamir, P. Research on teaching in the natural sciences. In R.M.W. Travers (Ed.), Second handbook of research on teaching. Chicago: Rand McNally, 1973, pp. 1098-1148.

Shumsky, A. In search of teaching style. New York: Appleton-Century-Crofts, 1968.

Siegel, L. The instructional gestalt: A conceptual framework. Teachers College Record, 1960, 62, 202-213.

Siegel, L., & Siegel, L.C. Educational set: a determinant of acquisition. Journal of Educational Psychology, 1965, 56, 1-12.

Siegel, S. Non-parametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

Simon, A., & Boyer, E.G. (Eds.) Mirrors for behavior: An anthology of classroom observation instruments. Philadelphia: Research for Better Schools Inc., 1967.

Smith, B.O. A concept of teaching. Teachers College Record, 1960, 61, 229-241.

Smith, B.O., & Meux, M. A study of the logic of teaching. In R.T. Hyman (Ed.), Teaching: Vantage points for study. New York: Lippincott, 1968, pp.101-117. (Reprinted from USOE Cooperative Research Project No. 258 (7257). Urbana: Bureau of Educational Research, College of Education, University of Illinois, 1962.)

Snow, R.E. Brunswikian approaches to research on teaching. American Educational Research Journal, 1968, 5, 475-489.

Snow, R.E. Representative and quasi-representative designs for research on teaching. Review of Educational Research, 1974, 44, 265-291.

Snow, R., Tiffin, J., & Siebert, W.F. Individual differences and instructional film effects. Journal of Educational Psychology, 1965, 56, 315-326.

Soar, R.S. New developments in effective teaching. The American Biology Teacher, 1968, 30, 43-47.

Sparrow, N.H., & Ross, J. The dual nature of extraversion: A replication. Australian Journal of Psychology, 1964, 16 214-218.

Stephenson, W. The inverted factor technique. British Journal of Psychology, 1936, 26, 344-361.

Stephenson, W. The study of behavior: Q-technique and its methodology. Chicago: University of Chicago Press, 1953.

Stern, G.S. Measuring non-cognitive variables in research on teaching. In N.L. Gage (Ed.), Handbook of Research on Teaching. Chicago: Rand McNally, 1963, pp.398-447.

Strike, K.A. The logic of learning by discovery. Review of Educational Research, 1975, 45, 461-483.

Suchman, J.R. Inquiry training in the elementary school. The Science Teacher, 1960, 27, 42-47.

Suchman, J.R. The child and the inquiry process. In Intellectual development: Another look. Washington, D.C.: Association for Supervision and Curriculum Development, 1964, pp.59-77.

Swaminathan, H., Hambleton, R.K., & Algina, J. Bayesian decision - theoretic procedure for use with criterion-referenced tests. Journal of Educational Measurement, 1975, 12, 87-98.

Taba, H., & Elzey, F.F. Teaching strategies and thought processes. Teachers College Record, 1964, 65, 524-534.

Taba, H., Levine, S., & Elzey, F. Thinking in elementary school children. USOE Cooperative Research Project No. 1574. San Francisco: San Francisco State College, 1964.

Taft, R. Test review. Omnibus Personality Inventory (OPI) Form F by P. Heist and G. Yonge and California Psychological Inventory (CPI) by H.C. Gough. Bulletin for Psychologists (Australian Council for Educational Research), August, 1969, No. 11, 2-5.

Tallmadge, G.K., & Shearer, J.W. Relationships among learning styles, instructional methods and the nature of learning experiences. Journal of Educational Psychology, 1969, 60, 222-230.

Tallmadge, G.K., & Shearer, J.W. Interactive relationships among learner characteristics, types of learning, instructional methods, and subject matter variables. Journal of Educational Psychology, 1971, 62, 31-38.

Tamir, P. An inquiry oriented laboratory examination. Journal of Educational Measurement, 1974, 11, 25-33.

Tamir, P. The relationship among cognitive preference, school environment, teachers' curricular bias, curriculum, and subject matter. American Educational Research Journal, 1975, 12, 235-264.

Tamir, P., & Glassman, F. A practical examination for BSCS students. Journal of Research in Science Teaching, 1970, 7, 107-112.

Tamir, P., & Glassman, F. A laboratory test for BSCS students - A progress report. Journal of Research in Science Teaching, 1971, 8, 332-341. (a)

Tamir, P. & Glassman, F. Laboratory test for BSCS students. BSCS Newsletter, 1971, No. 42, 9-13. (b)

Tatsuoka, M.M. Multivariate analysis in educational research. In F.N. Kerlinger (Ed.), Review of research in education, 1. Itasca, Ill.: Peacock, 1973, pp.273-319.

Taylor, J.A. A personality scale of manifest anxiety. Journal of Abnormal and Social Psychology, 1953, 48, 285-290.

Thomas, B. & Snider, W. The effects of instructional method upon the acquisition of inquiry skills. Journal of Research in Science Teaching, 1969, 6, 377-386.

Thurstone, L.L. Multiple factor analysis. Chicago: University of Chicago Press, 1947.

Tisher, R.P. Practical insights gained from Australian research on teaching. The Australian Science Teachers Journal, 1977, 23(2), 99-104.

Tisher, R.P., & Power, C.N. A study of the effects of teaching strategies in ASEP classrooms. Australian Journal of Education, 1975, 19, 127-145.

Tobias, S. Achievement treatment interactions. Review of Educational Research, 1976, 46, 61-74.

Tobias, S., & Ingber, T. Achievement-treatment interactions in programmed instruction. Journal of Educational Psychology, 1976, 68, 43-47.

Trebilco, G.R. A study of changes in classroom behaviour and of attitudes in student teachers trained in interaction analysis. Unpublished M.Ed. thesis, Monash University, 1975.

Tuckman, B.W. A technique for the assessment of teacher directiveness. Journal of Educational Research, 1970, 63, 395-400.

Turner, R.L. Conceptual foundations of research in teacher education. In B.O. Smith (Ed.), Research in teacher education. Englewood Cliffs, N.J.: Prentice-Hall, 1971, pp.10-36.

Tyler, L.E. Individual differences. In R.L. Ebel (Ed.), Encyclopaedia of educational research. New York: Macmillan, 1969, pp.639-645.

Veldman, D.J. Fortran programming for the behavioral sciences. New York: Holt, Rinehart, & Winston, 1967.

Walters, L., & Sieben, G. Cognitive style and learning science in elementary schools. Science Education, 1974, 58, 65-74.

Wang, M.C., Resnick, L.B., & Boozer, R.F. The sequence of development of some early mathematics behaviors. Child Development, 1971, 42, 1767-1778.

Ward, J.H., Jnr. Hierarchical grouping to maximize an objective function. Journal of the American Statistical Association, 1963, 58, 236-244.

Weick, K.E. Systematic observational methods. In G. Lindzey & E. Aronson. (Eds.), Handbook of social psychology (rev. ed.). Reading, Mass.: Addison-Wesley, 1968.

Weir, E.C. The open mind: An essential in teaching and learning. The Educational Forum, 1963, 27, 429-435.

West, L.H.T. Comment on 'The proper experimental unit: Comparative analysis of empirical data.' Journal of Research in Science Teaching, 1977, 14, 163-166.

White, R.T. Some aspects of methods for placing people in groups. Science Education: Research 1973, Report of the Australian Science Education Research Association Conference, 1973, 163-167.

White, R.T. The validation of a learning hierarchy. American Educational Research Journal, 1974, 11, 121-136.

Williams, C. A study of cognitive preferences. Journal of Experimental Education, 1975, 43(3), 61-77.

Wish, M. An unfolding analysis of cognitive preferences. Research Bulletin RB-64-7. Educational Testing Service, Princeton, 1974.

Wispé, L.G. Evaluating section teaching methods in the introductory course. Journal of Educational Research, 1951, 45, 161-186.

Withall, J. The development of a technique for the measurement of social-emotional climate in classroom. Journal of Experimental Education, 1949, 17, 347-361.

Witkin, H.A., Dyk, R.B., Faterson, H.F., Goodenough, D.R., & Karp, S.A. Psychological differentiation. New York: Wiley, 1962.

Witkin, H.A., Goodenough, D.R., & Karp, S.A. Stability of cognitive style from childhood to young adulthood. Journal of Personality and Social Psychology, 1967, 7, 291-300.

Witkin, H.A., Moore, C.A., Goodenough, D.R., & Cox, P.W. Field-dependent and field-independent cognitive styles and their educational implications. Review of Educational Research, 1977, 47, 1-64.

Wittrock, M.C. Response mode in the programming of kinetic molecular theory concepts. Journal of Educational Psychology, 1963, 54, 89-93.

Wittrock, M.C. The learning by discovery hypothesis. In L.S. Shulman & E.R. Keislar (Eds.), Learning by discovery: A critical appraisal. Chicago: Rand McNally, 1966, pp.33-75.

Wright, C.J., & Nuthall, G. Relationships between teacher behaviors and pupil achievement in three experimental elementary science lessons. American Educational Research Journal, 1970, 7, 477-491.

Wright, E.M.J. Development of an instrument for studying verbal behaviors in a secondary school mathematics classroom. Journal of Experimental Education, 1959, 28(2), 103-121.

Yager, R.E. Comparison of teaching outcomes between two teachers of secondary biology. The American Biology Teacher, 1968, 30, 816-819.

Yager, R.E., Engen, H.B., & Snider, B.C.F. Effects of the laboratory and demonstration methods upon the outcomes of instruction in secondary biology. Journal of Research in Science Teaching, 1969, 6, 76-86.

Yando, R.M., & Kagan, J. The effect of teacher tempo on the child. Child Development, 1968, 39, 27-34.

Yerkes, R.M., & Dodson, J.D. The relation of strength of stimulus to rapidity of habit-formation. Journal of Comparative Neurology and Psychology, 1908, 18, 459-482.